

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2004-500614

(P2004-500614A)

(43) 公表日 平成16年1月8日(2004.1.8)

(51) Int. Cl. <sup>7</sup>

F 1

テーマコード (参考)

G 0 6 F 17/30

G 0 6 F 17/30

1 7 O F

2 G 0 4 5

C 1 2 M 1/00

C 1 2 M 1/00

A

4 B 0 2 4

G 0 1 N 33/15

G 0 1 N 33/15

Z

4 B 0 2 9

G 0 1 N 33/48

G 0 1 N 33/48

Z

5 B 0 7 5

G 0 1 N 33/50

G 0 1 N 33/50

Z

審査請求 未請求 予備審査請求 有 (全 57 頁) 最終頁に続く

(21) 出願番号 特願2000-614100 (P2000-614100)  
 (86) (22) 出願日 平成12年4月26日 (2000.4.26)  
 (85) 翻訳文提出日 平成13年10月24日 (2001.10.24)  
 (86) 国際出願番号 PCT/US2000/011073  
 (87) 国際公開番号 WO2000/065421  
 (87) 国際公開日 平成12年11月2日 (2000.11.2)  
 (31) 優先権主張番号 60/130,992  
 (32) 優先日 平成11年4月26日 (1999.4.26)  
 (33) 優先権主張国 米国 (US)

(71) 出願人 501414412  
 ノヴァスクリーン バイオサイエンス  
 コーポレーション  
 アメリカ合衆国 21076 メリーラン  
 ド州, ハンオーバー, スタンダード ドラ  
 イブ 7170  
 (74) 代理人 100091096  
 弁理士 平木 祐輔  
 (74) 代理人 100102576  
 弁理士 渡辺 敬章  
 (74) 代理人 100103931  
 弁理士 関口 鶴彦

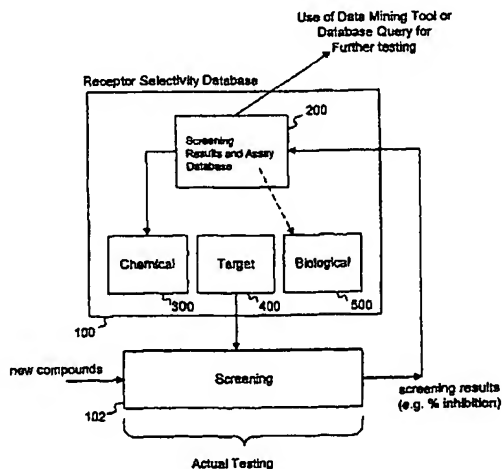
最終頁に続く

(54) 【発明の名称】 レセプタ選択性マッピング

## (57) 【要約】

【解決手段】 化合物に関する記録及び該化合物の人間及び動物の生物学的システム上の作用に関する記録を含む第1データベースと、分子ターゲットに関する記録を含む第2データベースとからなるコンピュータシステム。該システムはさらに、第1データベースの化合物と第2データベースの分子ターゲットとの間の結合、反応性その他の相互作用テストに関する記録を含む第3データベースを含む。上記テストには、第2データベース中の特定分子ターゲット及びこれと相互作用することが知られた化合物間の相互作用において、前記化合物が有する作用効果に関する情報が含まれている。前記テストの閾値を設定する手段、及び前記相互作用テスト閾値を満たす化合物を選択する手段もまた、前記システムに含まれている。第1、第2及び第3データベースの情報を、ユーザが閲覧、操作又は分析することができるユーザインタフェースが設けられている。

【選択図】 図 5



## 【特許請求の範囲】

## 【請求項 1】

複数の化合物に関連する記録と、該化合物の生物学的システムに対する作用に関する生物学的情報に関連する記録とを含む第 1 のデータベースと、

複数の分子ターゲットに関連する記録を含む第 2 のデータベースと、

第 1 のデータベースの化合物と第 2 のデータベースの分子ターゲットとの間の相互作用のテストであって、前記複数の分子ターゲットと相互作用することが知られている化合物及び前記分子ターゲットの間の相互作用に対して、前記複数の化合物中の 1 化合物が有する作用の情報を含むテストに関連する記録を含む第 3 のデータベースと、

ユーザが、選択された化合物を閲覧し、第 1 のデータベースの化合物データの記録、又は第 2 のデータベースの分子ターゲットの記録との関連に応じて、第 1 のデータベース、第 2 のデータベース及び第 3 のデータベースから選択的に情報を閲覧することができるようになっているユーザインタフェースと、  
からなるコンピュータシステム。

10

## 【請求項 2】

前記相互作用には結合が含まれ、前記作用には阻害作用が含まれることを特徴とする請求項 1 に記載のコンピュータシステム。

## 【請求項 3】

前記化合物には、知られた生物活性がない化合物、又は試験をパスしなかった化合物が含まれることを特徴とする請求項 1 に記載のコンピュータシステム。

20

## 【請求項 4】

前記化合物には、動物でテストされた化合物が含まれることを特徴とする請求項 1 に記載のコンピュータシステム。

## 【請求項 5】

前記化合物には、環境に作用することが知られている化合物が含まれることを特徴とする請求項 1 に記載のコンピュータシステム。

## 【請求項 6】

前記化合物には、薬理学的参照剤が含まれることを特徴とする請求項 1 に記載のコンピュータシステム。

## 【請求項 7】

前記化合物には、臨床用薬品マーケットで知られており、利用可能な生物学的情報が相当量存在する薬物が含まれることを特徴とする請求項 1 に記載のコンピュータシステム。

30

## 【請求項 8】

前記化合物には、人間でのテストを認可された化合物が含まれることを特徴とする請求項 1 に記載のコンピュータシステム。

## 【請求項 9】

前記化合物には、自然資源から得られる生物活性を示す化合物が含まれることを特徴とする請求項 1 に記載のコンピュータシステム。

## 【請求項 10】

前記分子ターゲットには、レセプタが含まれることを特徴とする請求項 1 に記載のコンピュータシステム。

40

## 【請求項 11】

前記分子ターゲットには、酵素が含まれることを特徴とする請求項 1 に記載のコンピュータシステム。

## 【請求項 12】

前記分子ターゲットには、核酸が含まれることを特徴とする請求項 1 に記載のコンピュータシステム。

## 【請求項 13】

前記分子ターゲットには、炭水化物が含まれることを特徴とする請求項 1 に記載のコンピュータシステム。

50

## 【請求項 1 4】

複数の化合物に関連する前記第 1 のデータベースの記録は、前記化合物の記述と特性に関するカテゴリーによって体系化されていることを特徴とする請求項 1 に記載のコンピュータシステム。

## 【請求項 1 5】

前記カテゴリーには、化合物名、化合物のタイプ、物理化学的特性、化学的空間配置又は化学的構造のデスク립タ、及び可溶性が含まれることを特徴とする請求項 1 4 に記載のコンピュータシステム。

## 【請求項 1 6】

前記第 1 のデータベースには、自然物質のデータベースが含まれることを特徴とする請求項 1 に記載のコンピュータシステム。

10

## 【請求項 1 7】

前記第 1 のデータベースには、テストをパスしなかった薬物のデータベースが含まれることを特徴とする請求項 1 に記載のコンピュータシステム。

## 【請求項 1 8】

前記第 1 のデータベースには、化学レジストリのデータベースが含まれることを特徴とする請求項 1 に記載のコンピュータシステム。

## 【請求項 1 9】

前記第 2 のデータベースには、3次元構造のデータベースが含まれることを特徴とする請求項 1 に記載のコンピュータシステム。

20

## 【請求項 2 0】

前記第 2 のデータベースには、シーケンス／突然変異のデータベースが含まれることを特徴とする請求項 1 に記載のコンピュータシステム。

## 【請求項 2 1】

前記第 2 のデータベースには、遺伝子データベースが含まれることを特徴とする請求項 1 に記載のコンピュータシステム。

## 【請求項 2 2】

前記化合物が前記生物学的ターゲットに対して有する作用に関する生物学的情報に関連する前記第 3 のデータベースの記録は、化合物名、ターゲット名、毒性、副作用、及び薬理作用のメカニズムを含むカテゴリーにより体系化されることを特徴とする請求項 1 に記載のコンピュータシステム。

30

## 【請求項 2 3】

前記作用に関連する相互作用テスト閾値を設定するための手段と、前記化合物を使用すると前記相互作用テスト閾値を満たす結果が得られる場合に、前記化合物を選択するための手段とをさらに含む請求項 1 に記載のコンピュータシステム。

## 【請求項 2 4】

複数の化合物に関連する記録を含む第 1 のデータベースから化合物を選択するステップと、

複数の分子ターゲットに関連する記録を含む第 2 のデータベースから分子ターゲットを選択するステップと、

40

前記選択された化合物のそれぞれと前記選択された分子ターゲットのそれぞれとの間の相互作用に関連する情報を生成するステップと、

化合物の生物学的ターゲットに対する作用に関する生物学的情報に関連する記録を含む第 3 のデータベースから生物活性を選択するステップと、

前記生成された情報を用いて、化合物及び分子ターゲット間の相互作用のパターンを、前記選択された生物活性と関連付けるステップとからなる新薬発見及び開発に関連するデータを分析するための方法。

## 【請求項 2 5】

前記情報を生成するステップは、

前記選択された化合物のそれぞれと前記選択された分子ターゲットのそれぞれとの間の結

50

合性に関する結合データを、未知の化合物が該結合に対して有する阻害作用をモニタすることにより、生成するステップと、  
前記阻害作用に関する結合テストの閾値を設定するステップと、  
未知の化合物、分子ターゲット、及び前記結合テストの閾値を満たす又は満たさない化合物の組み合わせに関する情報を生成するステップとを含むことを特徴とする請求項 2 4 に記載の方法。

【請求項 2 6】

前記結合性データは、ポジティブ及びネガティブの結合性情報を含むことを特徴とする請求項 2 5 に記載の方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、一般的に、化学インフォマティクス及びバイオインフォマティクスと、化学分子ターゲットの相互作用に関するデータとの組み合わせにより、多次元データベースを構築する技術に関する。特に、本発明は、化合物、分子ターゲット、及び生物学的又は臨床的情報を含むデータベースに関するものであり、該データベースにおいて、化合物及び分子ターゲット間の相互作用のパターン又は関係性が決定され、データベース内の他の情報と比較されることにより、薬物の発見及び開発、並びに関連領域において有用となる結論が得られるようになっている。

【0002】

【従来の技術】

世界規模の製薬業界では、研究開発に年間 300 億ドルが費やされており、このうちおよそ 3 分の 1 は、前臨床開発及び臨床開発のための薬物候補を選択するまでの期間である発見段階及び初期開発段階に費やされている。薬物発見に重要な臨床的ステップは、以下のステップからなる。(1) ヒトゲノムのセグメントを含む DNA のシーケンス、(2) 特定の病気又は生物学的機能と関連するゲノムを有する遺伝子の同定、(3) 該機能的遺伝子に関連し、又は該機能的遺伝子によりエンコードされ、後に薬物発見のための生物学的ターゲット又は分子ターゲットとなるレセプタ又は酵素などのタンパク質の生成、(4) 分子ターゲットに対して活性を化合物ライブラリーからスクリーニング、(5) 他の生物学的ターゲットに対して最も活性の高い化合物をスクリーニングして、目的の生物学的／分子ターゲットに対する該化合物の選択性又は特異性について評価し、また他のターゲットに対する活性により望ましくない副作用が起こる可能性について評価、(6) 毒性、吸収性、分布、代謝、排出等の特性を検定するための一定範囲のアッセイにおいて、最も作用が強く選択性の高い化合物を評定、(7) 上記の情報をを用いた経験的判断に基づいて最も有望な化合物を評定し、該情報を化学合成グループに送り初期活性化合物の類似体を生成、(8) 該化学的類似体をステップ(4)、(5)及び(6)において再テスト後、最適な誘導化合物又は化合物群が同定されるまでステップ(7)を繰り返す、(9) この最適な誘導化合物を更なる前臨床試験及び臨床試験に使用する。

【0003】

この発見及び開発のプロセスにおいて、狭いフィルターを通過した化合物は、より高価な前臨床開発及び臨床開発ようとして選別される。残念ながら、この選別プロセスに続く前臨床開発及び臨床開発において、化合物がこれらの段階をクリアできず、商品化にまでたどり着かないことがしばしばある。これらの失敗のため、1つの新薬を開発し発売するための平均コストは 3 億ドルを超えると推定されている。しかしながら、もし、発見及び開発プロセスの初期段階で最適な薬物候補を正確に同定することができ、該薬物が前臨床試験及び臨床試験をクリアすることができれば、開発コストは 75%も低減することができる。明らかに、製薬の研究開発(R & D)における主要目的は、上記のような薬物開発初期段階のテストの予測可能性を向上させることにある。

【0004】

バイオテクノロジーにおける技術革新と、実験プロセスの多くを自動化できる器具の発達

10

20

30

40

50

により、製薬 R & D に重大な影響を与える 2 つの主要なトレンドが生まれている。第一に、ヒトゲノムのシーケンスにおける進歩により、新薬発見スクリーニングプログラムに使用できる（新規レセプタ及び酵素などの）分子ターゲットの数は、急激に増加し続けている。約 400 個の分子ターゲットが新薬発見のために調査されており、ヒトゲノム計画により解明されるであろう潜在的分子ターゲットの数は数千から 1 万以上と推測されている。第二に、自動化及び組み合わせ化学（c o m b i n a t i o n a l c h e m i s t r y）などの新技術により、新薬発見スクリーニングプログラムに用いることができる化合物ライブラリーのサイズは、およそ 10 倍（多くの製薬企業が有する 100 万種以上の化合物）に膨張している。これらの 2 つの要因は新薬発見に大きな見込みを与えるが、一方では、新薬開発のコスト面において望ましくない結果をもたらす重大な潜在的問題をも生み出している。さらに多くのターゲット及び化合物が更なる生物活性化合物の発見につながり、その結果、前臨床試験に進むための最適な薬物候補の選択において大きな困難が生まれ、また、より多くの化合物が前臨床試験及び臨床試験に進むことによりこれらの段階でより多くの失敗が生まれるため、開発コストは増大するであろう。

#### 【0005】

これらの要因により、誘導化合物の選別、最適化及び検認においては、迅速かつ低コストな（「試験管」又はマイクロプレートベースの）インビトロ・アッセイがますます必要となる。このような迅速なアッセイは、後の高価な新薬開発段階に移行する前に、これらの活性化合物のうち最も有望なものを同定するのに役立つであろう。これらの要因により、さらに、遺伝子及び遺伝子生成物（分子ターゲット）、化学的構造、及びスクリーニング結果に関する膨大なデータを管理し解釈するためのより効果的な方法が必要となる。

#### 【0006】

製薬 R & D において重要性が増しているインビトロ・アッセイの応用の一つは、「プロファイリング」である。この特許出願の権利者は、1980 年代後半にプロファイリングの概念を開拓した。製薬企業は、新薬として開発される化合物の薬物学的活性及び潜在的副作用の特徴付けのためのインビトロ・アッセイの膨大なアレイを有している。現在、中枢神経系障害、免疫病、痛み及び炎症、感染症、癌、代謝又は成長因子、心臓血管機能、及び内分泌系に関わる病気を含む、広範囲の人間の病気において重要な役割を果たす分子ターゲット、必要なレセプタ及び酵素に基づいて、日常的に行われる 200 以上の異なるアッセイが存在する。薬品は、細胞レセプタとの相互作用により、世界市場機能の半分以上をしめている。さらに、多くの薬品の副作用は、そのレセプタ及び酵素との相互作用により緩和されるようになっている。

#### 【0007】

プロファイリングにおいて、ある製薬企業の、一般的には前臨床開発段階に入っている誘導化合物は、レセプタ及び酵素アッセイ装置によりテストされる。プロファイリングプロセスで得られる、この企業の化合物と特定のレセプタとの相互作用に関する情報は、誘導化合物の最適化及び選別において重要であり、また化合物の副作用又は第 2 の効能の可能性についての示唆ともなる。この知識により、この製薬企業は、該化合物の前臨床及び／又は臨床開発にかかる時間と費用数百万ドル分を潜在的に節約することができる。

#### 【0008】

長年、プロファイリングサービスが行われてきたが、製薬企業は一般的に、これらのテストから得られるデータを経験則的に使用していた。多くの薬物は、選択性の高い薬物も含めて、数多くのレセプタ又は他の分子ターゲットと相互作用する。したがって、プロファイリングにより生成されるデータの解釈は、製薬企業の研究者が、経験及び知識に基づき、化合物の化学的構造及び化合物と特定のレセプタとの結合作用の両方のデータを参酌して行うものである。残念ながら、最も経験を積んだ薬理学者でさえ、様々な薬物と、新薬開発に関連した広範囲のレセプタとの相互作用についての知識は完全ではない。

#### 【0009】

遺伝子及び遺伝子生成物（分子ターゲット）、化学的構造、及びスクリーニング結果に関する膨大なデータを管理し、照合し、解釈し、及び活用するためのより効果的な方法の必

10

20

30

40

50

要性から、バイオ情報学及び化学情報学、又は生物学的及び科学的データの管理における新たな機会が創造されている。新薬発見のための情報の膨大なプールを作り出す諸段階は、以下の各段階からなる。(1) DNAシーケンス(細胞が遺伝子生成物又はタンパク質を生成するための設計図となる遺伝的物質又は遺伝子のコーディング)、(2) 機能的ゲノミクス(特に薬物又は生物学的機能の変化に応じて、mRNA生成物を介して、DNAシーケンスを、関連する遺伝子生成物又はタンパク質に転換するプロセス)、(3) プロテオミクス(アミノ酸シーケンス、及び/又は、遺伝子にコーディングされているレセプタなどの遺伝子生成物又はタンパク質の3次元構造の同定)、(4) 微量分子の薬理学/毒物学(レセプタなどの遺伝子生成物と、薬物となり得る微小有機化合物との分子結合性又は相互作用)、(5) 化学的構造(微小分子、薬物類似化合物について)。

10

#### 【0010】

DNAシーケンスのためのデータベース(グループ1)は確立されており、Genbank、ゲノムセンターなどを含む。同様に、化学的構造のデータベース(グループ5)もまたよく知られており、MDL(Info)やOxford Molecularなどのベンダーにより提供されている。プロテオミクスのデータベース(グループ3)、例えばSWISS-PROT、ProLink及びPDBなども、また、構築されている。これらのデータベースは、構造情報を含んでおり、その1次元において、あるいは構造情報又はシーケンス情報の1コンポーネントにおいてパターンを決定するのに用いることができるので、それぞれを1個のコンポーネントとして考えることができる。グループ2及び4のデータベースはまだよく構築されていないが、新薬発見及び開発のための情報プールの貴重な追加情報となるだろう。これら後者2つの形式のデータベースは、遺伝子対タンパク質(グループ2)及びタンパク質対化合物(グループ4)などのように、2つの構造間の相互作用に関するデータを含んでいるので、2個の構成部品からなり、2次元である。このようなデータベースの関係性は、1構成部品からなるデータベースに比べて、複雑さのレベルが追加されたものである。

20

#### 【0011】

グループ4のタンパク質対化合物の関係性のための部分的データベース又は複数のデータベースが現在構築されている。例えば、本願権利者がクライアントに提供している、広範囲のレセプタターゲットに対する単一化合物の結合プロファイルは、グループ4タイプのデータベースの部分的データセットである。同様に、化学的構造データベース(グループ5)に含まれているような数千から数十万種の化合物を、特定のレセプタターゲット(グループ3内の単一点)に対する活性によりスクリーニングするような高度処理スクリーニングプロジェクトによって生成されたデータは、グループ4のデータベースの一部となるであろう。このような部分的グループ4データベースは新薬発見及び開発において役に立つであろうが、これらには2つの主要な欠点がある。第一に、これらは、単一化合物又は限られた化合物セットの、一定範囲のレセプタ(プロファイル)又は1個のレセプタターゲットにおける多数の化合物に対する結合選択性などのような、特定の2個コンポーネントの分析に関するものである(高度処理スクリーニングプロジェクト)。いずれのケースにおいても、複数のレセプタターゲットと複数の化学的構造との間の統計学的相関関係を扱えるほどデータセットの幅は十分に大きくはない。第二に、重要なことであるが、これらの部分的データセットは、構造的新規性、すなわち特に新薬としてのポテンシャルに基づいて選択された化合物に関して生成されるものである。これらは新規な化合物であるから、動物又は人間の体内における活性についての生物学的情報は全く存在しない。したがって、このようなアプローチは、上記のように、プロファイルのデータを経験的に解釈しようとする薬理学者と同じ限界に苦しめられることになる。

30

40

#### 【0012】

##### 【発明が解決しようとする課題】

したがって、本発明の目的は、新薬発見及び開発に関連したデータ分析のためのシステム及び方法を提供することにより、上記の必要を満たすことである。多数の化合物の多数の分子ターゲットに対するテスト結果から得られるポジティブデータ及びネガティブデータ

50

を含む、全分類スクリーニングデータベースが提供される。化合物及び分子ターゲットの組み合わせ数は、統計学的手法又は他のデータマイニング方法の分野において通常の知識を有する者が、このスクリーニングデータベースと、関連する化合物データベース及び分子ターゲットデータベースとを用いることにより、どの化合物が臨床試験に適しており、安全かつ効果的な薬物となる高い見込みがあるかについて、信頼性の高い予測をすることができる程度に大きくなくてはならない。

#### 【0013】

##### 【課題を解決するための手段】

本発明は、特に、上記要求を満たすようなシステム及び方法をここに開示する。このシステムには、複数の化合物に関連する記録及び該複数の化合物の人間及び動物の生物学的システム上の作用に関連する記録を有する第1のデータベースと、複数の分子ターゲットに関連する記録を有する第2のデータベースとからなるコンピュータシステムが含まれる。このコンピュータシステムは、さらに、第1のデータベース中の化合物と第2のデータベース中の分子ターゲットとの間の結合性、反応性、その他の相互作用に関するテストに関連する記録を有する第3のデータベースを含む。上記テストには、第2のデータベース中の複数の分子ターゲットから選択された、特定の分子ターゲットと相互作用することが知られている化合物（例えば、対照剤又は対照基準）との相互作用において、第1のデータベース中の複数の化合物から選択された化合物が有する作用効果についての情報が含まれており、また上記テストは、第2のデータベース中の複数の分子ターゲットに対して行われるようになっている。前記副作用に関連する相互作用テストの閾値を設定するための手段、及び化合物、化合物セット、及び／又は作用効果テストの結果が前記相互作用テスト閾値に合致するときのこれら化合物に関する情報を選択するための手段もまた、前記コンピュータシステムに含まれている。第1のデータベース及び第2のデータベース中の情報、並びに、第1のデータベース中の1以上の化合物の記録と、及び／又は第2のデータベース中の1以上の分子ターゲットの記録とに関連する第3のデータベース中の情報を、特に前記相互作用テスト閾値に合致する結果に関連した化合物、分子ターゲット、その他のデータベース記録に関して、ユーザが閲覧し及び操作し、又は分析することができるよう、ユーザインタフェースが設けられている。

#### 【0014】

さらに、本発明は、統計学的手法及びその他のデータマイニング手法をこれらの多次元データベースに応用して、新薬発見及び開発に関連する相関関係又はパターンを決定することにも関連している。

#### 【0015】

上記の一般的記述及び以下の詳細な記述は、ともに、実施例と説明を提供するだけであり、これらの記述により本発明の特許請求の範囲が制限されるべきものではない。

#### 【0016】

本明細書の一部として含まれる添付図面は、本発明の実施形態を示すものであり、また、本明細書の記述と合わせて、本発明の利点と原理を説明するためのものである。

#### 【0017】

##### 【発明の実施の形態】

本発明の好ましい実施形態を説明するが、その実施例は添付図面に示されており、また本発明の詳細な説明からも明らかとなるであろう。異なる図面における同一の参照番号は、可能な限り、同一又は同類の要素を示すものとする。

#### 【0018】

本発明に一貫するシステム及び方法により、新薬発見及び開発に関連したデータを分析して、例えば、新規化合物が安全又は効果的な新薬となる可能性が高く、全臨床試験及び臨床試験の段階に進むべきかどうかを予測することが可能となる。以下の記載では、本発明のシステム及び方法を、複数のメインテーブルを含むリレーショナルデータベースに関連して、また、化合物及び分子ターゲット間の結合性をこれら2者間の相互作用の測定基準として使用する用法に関連して説明する。この説明はまた、一般的に、複数のメインコン

ポーネントを有する他のデータベース構造や、化合物及び分子ターゲット間の他の相互作用の測定にも適用され得るものである。

【0019】

本発明は、化合物、分子ターゲット、特にタンパク質、その他の高分子、及びこれら化合物の生物活性に関する情報豊富なデータベースの新規な設計、構造、及び応用に関するものである。本発明はさらに、全臨床試験又は臨床試験を通過しなかった既知の薬物及び薬物候補を、これらの化合物の副作用、反応機構その他の医学的データを含む前臨床データ及び臨床データとともに、データベースの化合物ライブラリーの情報源として使用する方法にも関する。本発明はまた、データベース中の化合物と分子ターゲットとの結合その他の相互作用を決定し、また、関係性分析法及びデータマイニング手法を用いて、これらの相互作用パターンと、新薬発見及び開発に関連した特定の生物反応との、又はこのような相互作用をする化合物の特定の化学的構造、下部構造、その他の特徴との、又はこのような相互作用をする分子ターゲットの生化学的特徴、構造、その他の特徴との相関関係を決定する。このようなデータマイニング技術の例として、以下を参照することができる。これらの参照資料は全て本発明に含まれるものとする。

a) Chen et al.、3次元デスクリプタを用いた大規模構造-活性データセットの再帰的分割分析 (Recursive Partitioning Analysis of a Large Structure-Activity Data Set Using Three-Dimensional Descriptors)、Journal of Chemical Information and Computer Sciences、1998年10月；

b) Hawkins et al.、再帰的分割を用いた大規模構造-活性データセットの分析 (Analysis of a Large Structure-Activity Data Set Using Recursive Partitioning)、Quant. Struct.-Act. Relat.、16:296-302 (1997年)；

c) DePriest et al.、アンギオテンシン転換酵素及びサーモリシン阻害剤の3D-QSAR；演繹的、経験的に決定される活性部位ジオメトリックスに基づくCoMFAモデルの比較 (3D-QSAR of angiotensin-converting enzyme and thermolysin inhibitors; a comparison of CoMFA models based on deduced and experimentally determined active site geometrics)、J. Am. Chem. Soc. 115:5372-84 (1993年)；

d) Good et al.、コンピュータ化学におけるレビュー (in Review in Computational Chemistry)；Lipkowitz, K. B.、Boyd, D. B. (des.)、VCH, New York, Vol. 7. pp 67-117 (1996年)；

e) Marshall et al. コンピュータ利用製薬設計 (Computer-Assisted Drug Design) ACS Symposium Series 112；American Chemistry Society: Washington, DC, 1979年；pp 205-226；

f) Molec et al.、3次元構造活性の関係性及び生物学的レセプタのマッピング (A three-dimensional structure activity of relationship and biological receptor mapping)、Mathematics and Computational Concepts in Chemistry；Ellis Horwood；Chichester, 1985年；pp 225-251；

g) Mayer et al.、構造活性研究に一致するアンギオテンシン転換酵素の活性部位の特異的ジオメトリ (A unique geometry of the active

10

20

30

40

50



- e site of angiotensin-converting enzyme consistent with structure activity studies) J. Comput. Aided Mol. Des., 1:3-16. (1987年);
- h) Sheridan et al., ディスタンスジオメトリ法に対するアンサンブルアプローチ (The ensemble approach to distance geometry): application to the nicotinic pharmacophore, J. med chem., 29: 899-906 (1986年);
- i) Martin et al., Pharmacophoreマッピングに対する高速で新しいアプローチ及びそのドーパミン及びベンゾジアゼピン拮抗薬への応用 (A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists) J. Comput. Aided Mol. Des., 7:83-102 (1993年); 10
- j) Catalyst/Hypo Tutorial, version 2.0, BioCAD Corp. Mountain View, CA, 1993年;
- k) Sprague, P. W., 化学的仮説の自動的生成及び触媒によるデータベース検索 (Automated chemical hypothesis generation and database searching with Catalyst), Perspect. Drug Discov. Des., 3:1-20 (1995年); 20
- l) Barnum et al. (Identification of common functional configurations among molecules), J. Chem. Inf. Comput. Sci., 1996, 36:563-71 (1996年);
- m) HipHop Tutorial, version 2.3; Molecular Simulation Inc.; Sunnyvale, CA, 1995年;
- n) Davies, K. 及び Upinn, R., 3D pharmacophore searching, net. Sci., (<http://www.org/Science/Cheminform/feature02.html>); 30
- o) Golender, V. 及び Vesterman, B., 薬物設計のためのAPEX 3Dエキスパートシステム (APEX 3D expert system for drug design), Net. Sci. (<http://awod.com/netsci/science/compchem/feature09.html>);
- p) Van Drie, J., pharmacophoreの3Dデータベースクエリー決定のための戦略 (Strategies for the determination of pharmacophoric 3D database queries), J. Comput. Aided Mol. Des., 11:39-52 (1997年);
- q) Van Dire, J. 及び Nugent, R., 組み合わせ化学が提起する問題への取り組み (Addressing the challenges posed by combination chemistry): 3D databases, pharmacophore; recognition and beyond, SAR QSA R Environ. Res., 9:1-21 (1998年); 40
- r) Finn et al., 帰納論理学プログラミングprogolを用いたPharmacophoreの発見 (Pharmacophore discovery using the inductive logic programming progol) Machine Learning, Special Issue on Applications and Knowledge Discovery, Kluwer Academic Publishers: Boston, 1998, pp 1-33; 50

s) Jain et al., 製薬設計のための形状に基づく機械学習ツール (Compass: a shape-based machine learning tool for drug design), J. Comput. Aided Mol. Des., 8: 635-52 (1994年)。

#### 【0020】

製薬業界における標準的操作手順とは違って、グループ4 データベースは2個より多いコンポーネントからなるデータベースとして構築されなければならない、また、レセプタ又は酵素ターゲットと化合物との両方において相当な範囲をカバーするものでなければならない、バックグラウンドセクションは示唆している。例として、3コンポーネントからなるデータベースを構築するには、まず、新薬発見及び開発に直接関連する情報を豊富に含んだ広範囲の化合物セットを選択する。最も関連性の高い情報は、人体については臨床試験及び／又は市販後調査において、動物については前臨床試験において、このような化合物をテストした実際の経験から得られることが多い。他の関連する生物学的情報は、1以上の生物活性を示す自然物質や、レセプタの生物学的特徴を研究する当業界で使用されている化学的参照基準から得ることができる。したがって、このようなグループ4のデータベースのために選択される情報豊富な化合物の一実施形態には、市販の薬品、臨床試験又は前臨床試験をクリアしなかった薬物、生物活性な自然物質又は自然抽出物、及びレセプタ結合アッセイに用いられる対照剤などが含まれる。

10

#### 【0021】

このようなデータベースは、科学文献から得られるスクリーニングデータを用いて構築することができる。このアプローチでは部分的データセットを生成することができるであろうが、限界がある。第一に、文献参照によって得られるのは、一的に、ポジティブな情報（例えば、特定の化合物と特定のレセプタとの結合の阻害についてのレポート）だけである。有用な情報の比較を行うためには、ネガティブデータもポジティブデータと同様に重要である。さらに、ポジティブ及びネガティブの両データが揃っていないデータセットには、ある種の統計学的分析は適用できないだろう。第二に、それぞれ個別の、1の化合物の1のレセプタに対する結合性データに関する記事の定量的レポートと、他の化合物の同レセプタに対する結合性データに関する別のレポートとでは、アッセイの行われる方法が異なるため、比較することができない。したがって、グループ4の3コンポーネントデータベースを生成するための一実施形態は、広範囲のレセプタ又は酵素ターゲットについて、広範囲の化合物アレイをスクリーニングすることにより、一貫性のある比較結果を得て、ポジティブ及びネガティブの両データを確保するものとなるであろう。

20

30

#### 【0022】

化合物コンポーネント：化合物ライブラリーの選択及び化合物データの包括

本発明は、薬学研究開発に関連した生物活性が知られている化合物を1個のコンポーネントとして含むデータベースに関する。生物活性に関する情報は、化合物のデータベース又はテーブルに含むことができる。

#### 【0023】

例えば、これらの情報豊富な化合物には以下のものが含まれる。

(a) 未知の化合物と、レセプタ又は酵素など特定の分子ターゲットとの間の相互作用又は分子結合を測定するための薬理学的対照剤又は対照基準である化合物。このような対照剤化合物の例には、試験化合物とレセプタ又は酵素を含む分子ターゲットとの間の結合作用を特徴付けるために用いる化合物が含まれる。他の対照剤には、Sigma Aldrich Corp. の1団体であるResearch Biochemicals Inc. (RBI) のカタログや、その他の当業界に知られた情報源から選択された化合物が含まれ得る。これらの薬理学的対照剤化合物は、事前にテストされるか、及び／又は薬品として市販されているか、あるいは生物活性の高い自然物質であるため、以下の3カテゴリーと重複する可能性がある。

40

(b) 現在又は過去に医療用として市販され、また相当量の生物学的情報が利用可能な既知の薬物である化合物。これらの化合物はよく知られており、連邦食品医薬品局 (FDA)

50

A) などの米国政府官庁より入手可能な刊行物や、また民間企業又は非営利団体が発行する刊行物に掲載されている。非営利団体から発行されているこのような刊行物の一つは、米国薬局方会議 (United States Pharmacopeial Convention Inc.) による USP DI Series であり、この中の、第 I 章、医療業者のための薬物情報 (Volume I. Drug Information for the Health Care Professional) は、USP DI Update により毎月更新されている。販売許可された新薬は、このカテゴリーに属することとなる。市販の薬品、又は FDA 又はこれと同等の外国の規制団体に認可された薬物は、公開された記録となるので、通常の知識を有する当業者はこのカテゴリーに属するような化合物を容易に同定することができるだろう。

10

(c) 新薬となる見込みがあるとして IND (治験薬: Investigational New Drug) ステータスを与えられたが、FDA からの認可を得るための臨床試験において十分な効能又は安全性を達成できなかった化合物、あるいは市販薬品としてのステータスを得るには至らなかった化合物などの、人体でのテストを許可された化合物。このカテゴリーの化合物には、FDA に販売認可されたが、後に市場から撤退したような化合物も含まれ得る。これらの化合物もまた、相当量の有用な生物学的情報を含んでおり、特に本発明の目的のために有用となるであろう。失敗した薬物の識別情報は、製薬企業又はバイオテクノロジー企業からの公式発表、“Pink Sheets” などの刊行物、及び FDA が保有しているリストなど、多くの情報源から得ることができる。

(d) 植物、微生物、動物等の自然資源から得られる生物活性を示す化合物。これらの自然物質には、新薬発見及び開発に関連する情報を与える、毒物、抗菌剤、行動調整剤 (behavioral modifier)、防衛剤、その他のカテゴリーの化合物が含まれるであろう。自然物質の識別情報は、RBI や Sigma Aldrich の化合物カタログなどの多くの刊行物から得ることができるが、これらに限られるわけではない。

20

#### 【0024】

このデータベースに含まれる化合物のそれぞれについて、化学的構造、化学式、物理化学的特性、化学的空間配置、その他の空間化学情報 (例えば、Smiles codes)、可溶性、その他の関連データが、利用可能な範囲で、データベースのフィールドに収録されている。通常の知識を有する者は、その他の収録可能なパラメータを認識することができるだろう。データベース中の化学構造の関係性から、又は他の関係性から、化合物を構築することができる。

30

#### 【0025】

図 1 A は、リレーショナルデータベースの化合物テーブル 300 を示す。テーブル 300 には複数の化合物がリストされており、複数の化合物 N の記録 (行 1 ~ N) を含んでいる。各化合物について、該化合物に関する情報を含む列 301 ~ 307 がある。例えば、図 1 A において、列 301 は化合物名を、列 302 は化合物の種類 (例えば、人体での試験を許可された化合物など) を、列 303 は化学的構造に関する情報、例えば、構造図を含むスクリーンを呼び出すためのハイパーリンク (図 1 B のスナップショット 310 を参照) を、列 304 は化合物の化学式を、列 305 は化合物の物理化学的特性に関する情報を、列 306 は化合物の空間配置を、列 307 は化合物の可溶性に関する情報を含んでいる。

40

#### 【0026】

テーブル 300 にリストされた各化合物 301 に関するその他の関連データを含めることができるように、さらに列を追加してもよい。これらの追加された列に化合物の生物活性を含めることにより、化合物データベースを 2 コンポーネントのデータベースにすることもできる (データベース 500 を参照)。

#### 【0027】

図 1 B は、テーブル 300 中の記録に関連する情報を含むスナップショット 310 を示す。例えば、化合物の化学式 304 を、化合物の構造 303 とともに、スナップショットに含めることができる。

50

## 【0028】

分子ターゲットコンポーネント：レセプタ、酵素、その他の分子ターゲットの選択及び分子ターゲットデータの包括

本発明のデータベースの第2のコンポーネントには、分子ターゲットとして、新薬発見及び開発に関連するレセプタ、酵素、その他のタンパク質、核酸、炭水化物、その他の高分子化合物などが含まれる。本発明の1実施形態では、レセプタ及び酵素が主要な分子ターゲットである。レセプタは、体内の細胞及び器官における分子レベルのコミュニケーションの大部分を緩和する。酵素は、例えば、2次的メッセンジャシステム及び細胞シグナル経路によって、このようなコミュニケーションを増大させることが多い。

## 【0029】

レセプタには、ドーパミンレセプタ、セロトニンレセプタ、アヘン剤レセプタ、ムスカリン性レセプタ、アドレナリン作用性レセプタ、アデノシンレセプタ等の典型的なレセプタ群が含まれる。これらのレセプタ群には、レセプタタイプのサブタイプ（ドーパミン-1、ドーパミン-2、ドーパミン-3、ドーパミン-4及びドーパミン-5レセプタなど）が含まれる。あるサブタイプは、さらなるバリエーション（ドーパミン4.2、ドーパミン4.4及びドーパミン4.7など）を有し、又は異なる形状（ドーパミン2 short及びドーパミン2 longなど）を持つものもある。特定のレセプタをエンコーディングする遺伝子の変異が起こることにより、薬物その他の化合物に対する結合性が通常のレセプタとはわずかに異なるレセプタ群のサブセットが誘導される可能性があり、レセプタのスプライス変形が出現する可能性もある。レセプタは、ファミリー、スーパーファミリー、又はサブファミリーにより分類することができる。Gタンパク質結合レセプタ、膜貫通型レセプタ、核レセプタ等の分類法がある。関連遺伝子のDNAシーケンスの類似性の程度によって、レセプタを分類することができる。また、アミノ酸シーケンス及びこれに関連する3次元配位によっても、レセプタを分類することができる。組織内のレセプタ出現位置によって、又は異なる細胞種にわたって、レセプタを分類することができる。

## 【0030】

酵素には、プロテアーゼ、カルボヒドラーゼ、キナーゼ、ホスホターゼ、DNA修飾酵素、トランスフェラーゼ、P450、その他の当業者に知られた酵素が含まれる。

## 【0031】

本出願人は、他のレセプタ、レセプタ供給源、及びこれに関するアッセイを常時開発して、データベースコンテンツに追加している。追加されるレセプタ及びレセプタアッセイは、当業者によく知られたものである。新薬発見及び開発に関連するレセプタについてのリスト及び記述は、当業者に知られている多くの刊行物から得ることができる。これらの刊行物には、RBIレセプタ分類ハンドブック（RBI Handbook of Receptor Classification）及びIUPHARレセプタ分類書（IUPHAR receptor classification book）が含まれる。さらに、新規のレセプタ及びレセプタサブタイプが発見されると、これらもデータベースコンテンツに追加される。

## 【0032】

酵素及び酵素アッセイは当業者によく知られている。新薬発見及び開発に関連するレセプタについてのリスト及び記述は、当業者に知られている多くの刊行物から得ることができる。

## 【0033】

図2は、分子ターゲット情報にアクセスするために使用できるリレーショナルデータベースシステムの部分を構成するテーブル400、410及び420を示す。テーブル400にはターゲットがリストされており、複数のターゲットMの記録（行1～M）を含んでいる。列401にはターゲット名がリストされており、列402には各ターゲット名についてターゲットのタイプが特定されている。

## 【0034】

テーブル構造は、列402に特定されるターゲットのタイプにより変化させてもよい。テ

10

20

30

40

50

ーブル410には、レセプタとして分類されテーブル400にリストされたターゲットに関する情報が含まれている。特定のレセプタ名についてデータベースをクエリーすることにより、テーブル410の記録にアクセスすることができる。テーブル410のレセプタ名は、列402の“レセプタ (Receptor)”と示されているターゲット名についてテーブル400をクエリーすることにより、アクセスが可能である。

#### 【0035】

テーブル410の列411にはレセプタ名が含まれているが、このレセプタ名はテーブル400の列401のターゲット名にもなっている。列412はレセプタのファミリー情報を含み、列413はレセプタのスーパーファミリー情報を含み、列414はレセプタのサブファミリー情報を含み、列415は関連遺伝子のDNAシーケンスの類似性の程度に関する情報を含み、列416はアミノ酸シーケンスに関する情報を含んでいる。アミノ酸シーケンスは、データベースに含まれている多数の分子デスクリプタの1つである。他の分子デスクリプタ417には、例えば、アミノ酸シーケンスに関連した水治療法プロットが含まれていてもよい。テーブル400、410及び420に示す分子ターゲットデータベースにはターゲット情報が含まれており、ターゲットに関連する生物学的情報もまたデータベースに含まれているため(テーブル600)、このデータベースは2コンポーネントからなると考えられる。ここに示した列は、データベースに含まれ得る情報のタイプを示すものであり、本発明を限定するものと解釈されるべきではない。

#### 【0036】

テーブル420には、テーブル400に酵素として分類されているターゲットの情報が含まれている。特定の酵素名についてデータベースをクエリーすることにより、テーブル420の記録にアクセスすることができる。テーブル420の酵素名は、列402に“酵素 (Enzyme)”と示されているターゲット名についてテーブル400をクエリーすることによりアクセスできる。

#### 【0037】

テーブル420の列421には酵素名が含まれているが、この酵素名はテーブル400の列401のターゲット名にも含まれている。列422は酵素のタイプに関する情報を含んでいる。列423は“その他の関連情報”とされており、ユーザがアミノ酸シーケンス及び分子デスクリプタを含む他の酵素情報にアクセスしたい場合などに応じて、さらなる列がテーブル420に追加される可能性があることを示している。

#### 【0038】

ターゲットのタイプによる分子ターゲット情報へのアクセスを説明するために、テーブル410及び420のみを示したが、このリレーショナルデータベースシステムには、データベースに使用可能な分子ターゲットターゲットのタイプの数に応じて、さらなるテーブルを追加することができる。

#### 【0039】

生物学的情報コンポーネント：生物学的／化学的情報のパラメータ

データベースの一部を構成する生物学的情報には、例えば、副作用、薬理作用のメカニズム、薬物による代謝、毒性、吸収性、分布及び排出などに関する事項が含まれている。これらの情報は、市販薬のFDA認可ラベルから、又は臨床試験をクリアしなかった薬物に関する文献及び刊行物から得ることができる。パラメータの具体例として、毒性、LD<sub>50</sub>、LD<sub>50</sub>/ED<sub>50</sub>、催奇性、毒性メカニズム、毒性のターゲットとなる器官、インビトロ毒性バッテリー、アポトーシス誘発性、生物学的利用性(バイオアベイラビリティ)、吸収性、血液脳関門、経口吸収性、粘膜吸収性、吸収率%、分布、限界血液タンパク質、半減期、作用の発現、作用の持続性、血液内濃度のピーク、代謝、主要経路、非主要経路、活性代謝物質、排出、第1排出様式、第2排出様式、インビトロ効用、治療法の指示、動物行動での作用、副作用、既知の主要なターゲット、その他のターゲット器官／システム、及び既知のレセプタ相互作用があげられる。

#### 【0040】

図3は、上記のうちいくつかの生物学的情報パラメータを含むテーブル500を示す。テ

10

20

30

40

50

ーブル500は、第1のデータベース内に存在し得る全ての化合物に関連するN個の行(1~N)からなる。列501には化合物名が含まれ、列502には(市販薬又は試験をクリアしなかった薬物についての)治療法の指示が含まれ、列503には毒性に関する情報が含まれ、列504には副作用に関する情報が含まれ、列505には薬物の作用メカニズムに関する情報が含まれている。例えば、テーブル500をテーブル300と関連付けることにより、化合物及び生物活性の2コンポーネントテーブルを構成することができる。

#### 【0041】

図3はまた、データベース内の分子ターゲットに関連した生物学的情報パラメータを含むテーブル600を示す。第2のデータベース内に存在し得る全てのターゲットに関連するテーブル600は、P個の行(1~P)からなる。列601にはターゲット名が含まれ、列602には(市販薬又は試験をクリアしなかった薬物についての)治療法の指示が含まれ、列603には毒性に関する情報が含まれ、列604には副作用に関する情報が含まれている。上記同様に、例えば、テーブル600をテーブル400と関連付けることにより、分子ターゲット及び生物活性の2コンポーネントテーブルを構成することができる。テーブル500及び600はともに、分子ターゲット情報、化合物情報、及び各分子ターゲット及び各化合物に関連した生物活性情報を含む全分類データベース(例えば、リレーショナルデータベースシステム内に存在し得る化合物及び分子ターゲットの全組み合わせを含むもの)とすることができ、また、多次元データベースとみなすことができる。本発明の範囲を逸脱することなく、テーブル500及び600に追加の列を加えることができる。

#### 【0042】

##### 結合性情報の決定

本発明の主要な特徴は、化合物、分子ターゲット及び生物学的情報からなる複数の情報コンポーネントの構成と、化合物及び分子ターゲット間の結合性、反応性、その他の相互作用についての評価である。この結合性又は反応性に関する情報を再度既知の生物学的情報に関連付けることにより、新薬発見及び開発に使用できるパターン又は関係性を選別することができる。本発明の重要な側面は、化合物及び分子ターゲット間の広範囲かつ一貫した結合性又は反応性データを生成することにより、必要なパターン又は関係性を特定するための可能な限り完全なデータセットを提供し、また、ポジティブ及びネガティブ両方の結合性又は反応性情報を提供することである。本発明の1実施例では、例えば、特定の分子ターゲット又は分子ターゲットセットに対して閾値セットを満たすか満たさないかを示す数値デスク립タとして、結合性データを構成する。この数値デスク립タは、生物学的システム又は生物学的情報セットに適する閾値にほぼ近い濃度において評価された、各化合物及び各レセプタ、その他の分子ターゲットに対する反応性の有無に関連付けることもできる。例えば、ある化合物について、レセプタ及びこれに対応する特異化合物間の結合阻害を、濃度 $10^{-5}$  M (10マイクロモル)、閾値30%でテストすることができる。この他に、初期濃度又は阻害率閾値を設定することができる。また、本発明の1実施形態においては、この最初のイエス/ノーテストにおいて閾値を超える結合阻害を起こす化合物について、結合阻害能をさらにテストする。これらの活性化合物について、例えば、 $10^{-5}$  ~  $10^{-9}$  Mの範囲内の7~14の異なる濃度を含む一連の濃度条件でテストを行い、特定のレセプタにおける活性化合物の $IC_{50}$ 、及び/又は $K_i$ の値を決定する。この決定をするための濃度条件の数はこれよりも多くとも少なくともよく、また、 $10^{-5}$  ~  $10^{-9}$  Mより高い又は低い濃度範囲を使用することが必要となることもあるだろう。これらのデータから、各分子ターゲットに対する各活性化合物の相対的活性度又は相対的効能のマトリクスが得られる。

#### 【0043】

これらのスクリーニングデータを生成するために、まず、化合物を適当な溶媒系において可溶化する。溶媒系は、例えば4% DMSOなどを用いることができるが、その他の濃度のDMSO又は他の溶媒を用いることもできる。次に、これらの化合物のストック溶液を適当な濃度に希釈して、貯蔵溶液(repository)として利用できるようにする

。化合物及び分子ターゲット間の相互作用を測定するための各アッセイごとに、用いる試薬及び手順は異なったものとなる。このような各アッセイを特徴づけ、ルーチン化して一貫性のあるものにすることが必要である。アッセイを行うごとに、適当な対照試験を行う必要がある。所望のタイプ及び正確さを持つ情報を生成することができるあらゆるアッセイフォーマットを使用することができる。放射性標識、蛍光分析、蛍光偏光分析、時間分解蛍光分析、蛍光相関性分光分析、化学ルミネセンス、UV吸光、比色分析など、数多くのアッセイ検出システムを使用することができる。

#### 【0044】

本発明の1実施例では、レセプタ結合アッセイ又は酵素活性アッセイを用いて分子相互作用に関するデータを生成する。例えば、レセプタ結合アッセイでは、貯蔵溶液の化合物を、レセプタと該レセプタ用に選択された参照剤との間の結合相互作用の阻害能についてテストする。レセプタは、動物又は人間の組織などから得ることができ、あるいは、レセプタ用の遺伝子を含むようトランスフェクションされた細胞株からも得ることができる。アッセイのレセプタ源を、例えば、レセプタを含む細胞フラクションとして用意することができる。レセプタは、また、部分的に精製されていてもよい。参照用の化合物又はリガンドは、特定のレセプタに対する潜在的な及び／又は特定の結合に基づいて選択されるのが好ましく、また、ヨウ素125、トリチウム、炭素14、その他の放射性トレーサを含むことにより、結合したリガンドと結合していないリガンドとを識別可能にすることができる。データベースに含まれるべき化合物の結合性データの試験と並行して、ポジティブ及びネガティブの対照試験を行い、参照（放射性）リガンドの様々な濃度での参照曲線により、行われたアッセイの質を保証する。

#### 【0045】

多数の方法及びシステムにより、ターゲット及び化合物間の相互作用を測定することができることを、当業者は理解できるだろう。結合反応を平衡状態に到達させるために、放射性リガンド、レセプタ調製溶液、及び試験化合物を、適当な時間、適当な緩衝溶液中、適当な温度下で培養する。結合放射性リガンドの非結合放射性リガンドに対する量は、ろ過や、SPA（scintillation proximity assay）などの方法を用いた分離ステップにより決定され、液体シンチレーション又はガンマ計数により測定される。次に、試験化合物のアッセイ結果とポジティブ及びネガティブ対照試験とを比較することにより、試験化合物の特定の結合数を決定する。これらのデータから試験化合物の阻害パーセントが計算される。

#### 【0046】

図4には、スクリーニング結果及びアッセイデータベースを表すテーブル200が示されており、データベース300（1～Nの化合物からなる）に含まれる化合物は、データベース400に含まれる分子ターゲットに対する作用についてテストされるようになっている。テーブル200は数多くの形式を取ることができる。例えば、テーブル210においては、複数の分子ターゲットの各々に対してテストされた複数の化合物の各々のスクリーニング結果が、各測定セットについて選択されたテスト結果閾値を超えるか下回るかに基づいて、スクリーニング結果を“イエス”又は“ノー”のエントリーとして入力することができる。

#### 【0047】

他の例では、スクリーニング結果は、複数の分子ターゲットの各々に対してテストされた複数の化合物の各々について、結合その他の作用の効能又は程度（例えば、化合物－レセプタ相互作用の $K_i$ ）を特定する数値デスク립タとして、テーブル220に入力される。好ましい実施形態では、テーブル210及び220において、このような“化合物”×“ターゲット”の全てのマトリクスポイントが決定され、全種類データベースが生成される。また、スクリーニング結果及びアッセイデータベース200は、他の化合物－ターゲット相互作用の測定結果を含んでいてもよく、この測定結果には、スクリーニング結果の生データ及びこの生データから得られる測定結果、アッセイの手順及び特徴、その他の関連情報が含まれる。

10

20

30

40

50

## 【0048】

図5A及び5Bは、ここではレセプタ選択性でとして例示してある、データベース100をスクリーニングプロセスの一部として使用することにより、更なる開発の対象となるべき新薬候補としての新規化合物を発見し選別すること（図5A）、又は、特定の病気検出のための新薬候補を発見するために使用する潜在的有効ターゲットとしての新規ターゲットを発見し選別すること（図5B）を示している。データベース100は化合物コンポーネント300、分子ターゲットコンポーネント400、生物学的情報コンポーネント500及び600、及びスクリーニング結果及びアッセイデータベース200を含むことができる。

## 【0049】

新規化合物又は化合物セットをスクリーニングプロセス102に導入して、これらが特定の化合物（例えば、参照剤）及び分子ターゲットに対する結合阻害能を有するかどうかを決定する（図5A参照）。スクリーニングプロセスでは、分子ターゲットコンポーネント400から得られるターゲット情報を使用することができる。

## 【0050】

スクリーニングプロセス102の結果は、中間データベースに記憶し、又はレセプタ選択性データベース100のスクリーニング結果及びアッセイデータベース200に入力することができる。また、この結果を特定のパラメータ（例えば、細胞毒性など）として、生物学的情報データベース500に記憶し、又は化合物データベース300に（例えば、化合物名として）記憶することもできる。

## 【0051】

スクリーニングプロセス102から得られる結果の完全なセットを、スクリーニング結果及びアッセイデータベース200に記憶することができる。化合物及び分子ターゲット間の結合に対して阻害能を示す新規化合物（例えば、参照剤）について、このデータベース200をクエリーすることにより、これらの新規化合物をさらにテストすることができる。

## 【0052】

これに代わって、例えば、「オーファン（orphan）」レセプタなどの新規分子ターゲットをスクリーニングプロセスに導入して、化合物データベース300の化合物に対するテストを行うこともできる（図5B参照）。オーファンレセプタについては、その構造は知られているが、機能及び病気との関連性は知られていない。新規分子ターゲットと相互作用する化合物の識別情報を含むスクリーニングプロセスの結果は、スクリーニング結果データベース200に組み入れられる。データベース100内をクエリーして、新規分子ターゲットの機能を特定し、及び／又は新規分子ターゲットの病気との関連性を確認する。

## 【0053】

図6Aは、データベース100を使用した新規化合物の薬物としてのポテンシャルの予想を示す。テーブル710は、化合物（300）、分子ターゲット（400）、生物学的情報（500及び600）、及びスクリーニング結果（200）のデータベースの情報に依存している。ユーザが新規化合物に関する情報をデータベース100に与えると、自動クエリースクリプトを実行してこの情報を取得することにより、テーブル710にはこれらのうち1以上のデータベース（又はテーブル）からの情報が含まれるようになっている。

## 【0054】

このテーブル710を生成するためのクエリースクリプトは、新規化合物の情報を与えられると、化合物を化合物データベース300から選択することができるようになっている。この選択は、新規化合物と、データベース300に既に含まれている化合物との間の、化学的構造その他の特性の類似性に基づいて行うことができる。

## 【0055】

化合物の選択後、クエリースクリプトは、選択された化合物と反応（結合）することが知られているターゲットをターゲットデータベース400から選択する。最後に、選択され

10

20

30

40

50



た化合物及び分子ターゲットを用いて、生物学的情報データベース500及び600をクエリーし、化合物-分子ターゲットの対に関する生物学的情報をテーブル710に挿入する。これに代わって、ユーザが興味のある特定の生物学的情報カテゴリー（例えば毒性）を入力することにより、テーブル710に含まれる生物学的情報をこのカテゴリーに限定することもできる。

【0056】

ユーザは、テーブル710をクエリーして、新規化合物の薬物として使用のポテンシャルの予測に関連した情報を得ることができる。この例としては、該新規化合物に関連した化合物と反応することが知られている分子ターゲットのクエリー、及びこの分子ターゲットが該化合物とともに引き起こすことが知られている副作用のクエリーが挙げられる。

10

【0057】

図6Bは、データベース100の使用をして、新規化合物の薬物ポテンシャルの予測と同様なアプローチにより、図6Bに示すデータ入力及びクエリーを用いて、新規分子ターゲットの病気との関連性及び／又は生物学的機能を確認する方法を示す。

【0058】

本明細書で言及した全ての特許、特許出願及び刊行物は、参照資料として本発明に含まれる。

【0059】

上記の本発明の実施形態の記述は、例示及び説明のためのものであり、網羅的なものではなく、また本発明をここに開示した形態に限定するものではない。上記の開示内容に照らして、又は本発明の実施に際して、本発明の修正や変更を行うことが可能であろう。

20

【図面の簡単な説明】

【図1】

図1Aは、本発明の1実施形態におけるレセプタ選択性マッピングデータベースの化合物テーブルを示す。

図1Bは、本発明の1実施形態におけるレセプタ選択性マッピングデータベースの化合物の空間配置を含む化合物記録のスナップショットを示す。

【図2】

本発明の1実施形態におけるレセプタ選択性マッピングデータベースの、分子ターゲット情報へのアクセスに使用可能な数個の論理テーブルを示す。

30

【図3】

本発明の1実施形態におけるレセプタ選択性マッピングデータベースの生物学的情報のテーブルを示す。

【図4】

本発明の1実施形態におけるレセプタ選択性マッピングデータベースをスクリーニングプロセスの一部として使用する方法を示す。

【図5】

図5Aは、レセプタ選択性マッピングデータベースをスクリーニングプロセスの一部として使用することにより、新薬候補となる新規化合物を発見し選別する方法を示す。

図5Bは、レセプタ選択性マッピングデータベースをスクリーニングプロセスの一部として使用することにより、特定の病気に対する新薬候補を発見するためのターゲット候補となる新規ターゲットを同定する方法を示す。

40

【図6】

図6Aは、新規化合物の薬物としての可能性を予測するためのデータベースの使用法を示す。

図6Bは、新規分子ターゲットの病気との関連性及び／又は生物学的機能を確認するためのデータベースの使用法を示す。

301	302	303	304	305	306	307
仕舞番号	タイプ	仕舞の用途	配方式	動機仕舞の条件	仕舞的 取引状況	可処分 現金
1						
2						
3						
4						
5						
6						
N						

A

【図2】

Figure 2 illustrates a data structure for enzyme and receptor information. The structure is organized into three main sections: 400, 410, and 420.

**Section 400 (Left):** A table with three columns: ターゲット名 (Target Name), ターゲットタイプ (Target Type), and レセプタ (Receptor). The rows are labeled 1, 2, and M.

**Section 410 (Middle):** A table with six columns: レセプタ (Receptor), サブファミリー (Subfamily), スーパーファミリー (Superfamily), DNS シーケンスの相対位置 (Relative position of DNS sequence), アミノ酸シーケンス (Amino acid sequence), and 構造 (Structure). The rows are labeled 1, 2, and M.

**Section 420 (Right):** A table with three columns: 酵素名 (Enzyme name), タイプ (Type), and 他の関連情報 (Other related information). The rows are labeled 1, 2, and M.

The diagram shows the following relationships:

- Section 400 is linked to Section 410 via arrows labeled 1 and 2.
- Section 410 is linked to Section 420 via arrows labeled 1 and 2.
- Section 420 is linked to Section 410 via arrows labeled 1 and 2.

生物学的／化学的指標テーブル

501	化合物名	502	503	504	505
	化合物1	精製法の精み	特性	製作用	発原作物の スカニズム
502	化合物2				
503	化合物N				

1 2 N 500

生物学的／化学的情報データベース

生物学的／化学的指標テーブル

501	502	503	504	505
化合物名	精製法の精査	特性	製作用 メカニズム	製薬会社の メカニズム
化合物 1				
化合物 2				
.				
化合物 N				

1

2

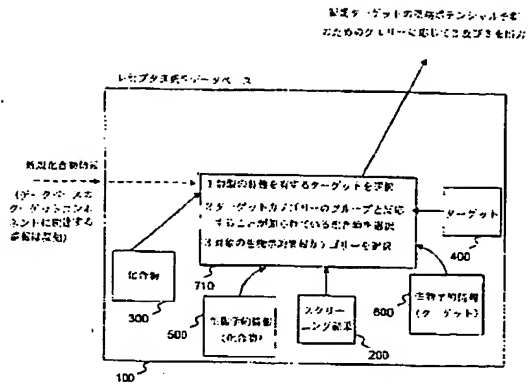
N

500

8


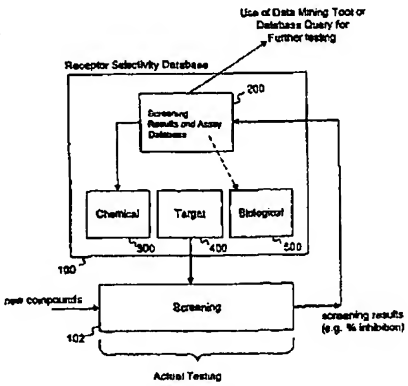
3105





**B**

## 【国際公開パンフレット】

<b>PCT</b> WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau		
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)		
(51) International Patent Classification: <b>G06F</b>	<b>A2</b>	(11) International Publication Number: <b>WO 00/65421</b>  (43) International Publication Date: <b>2 November 2000 (02.11.00)</b>
(21) International Application Number: <b>PCT/US99/11073</b>  (22) International Filing Date: <b>26 April 2000 (26.04.00)</b>  (30) Priority Data: <b>60/130,992</b> <b>26 April 1999 (26.04.99)</b> <b>US</b>  (71) Applicant: <b>OCEANIX BIOSCIENCES CORPORATION</b> <b>[US/US]: 7170 Standard Drive, Ilwaco, MA 21076 (US).</b>  (72) Inventor: <b>MANYAK, David, M.; (261) Folly Quarter Road, Elliot City, MD 21047 (US); ZEPPELLO, Renee, A.; 310 Ridgewood Road, Baltimore, MD 21210 (US); CHEN, Hao; 5905 Ohio Court, Columbia, MD 21044 (US); WEISSMAN, Arthur, D.; 3706 Manito Drive, Baltimore, MD 21215 (US); LANG, Gary, L.; 910 Southern Drive, Bel Air, MD 21014 (US).</b>  (74) Agents: <b>GARRETT, Arthur, S. et al.; Hienegun, Henderson, Farshaw, Garrett &amp; Dunner, L.L.P., 1300 I Street, NW, Washington, DC 20005-3315 (US).</b>		(81) Designated States: <b>AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GL, GR, GU, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, ND, NG, NL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AR/PTO patent (CH, GB, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), European patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, MG, MR, NE, NG, TD, TG).</b>  Published <i>Without international search report and to be republished upon receipt of that report.</i>
(54) Title: <b>RECEPTOR SELECTIVITY MAPPING</b>		
(57) Abstract  A computer system comprising a first database containing records corresponding to a plurality of chemical compounds and records corresponding to biological information related to effects of the plurality of chemical compounds on biological systems and a second database containing records corresponding to a plurality of molecular targets. The computer system further comprises a third database containing records corresponding to results of interaction between compounds in the first database and molecular targets in the second database, the records including information on the effect that a compound from the plurality of compounds has on the interaction of a compound known to interact with a molecular target from the plurality of molecular targets and said molecular target. Means for setting an interaction test threshold corresponding to said effect and means for selecting the compound when its test results in a test meeting the interaction test threshold are also included in the computer system. A user interface is provided to allow a user to view the selected compound and to selectively view information from the first database, the second database, the third database as it relates to a compound record in the first database or as it relates to a molecular target in the second database.		
		

## FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SP	Sweden
AN	Andorra	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Switzerland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GR	Greece	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	ML	The Gambia	TM	Turkmenistan
BF	Burkina Faso	GM	Ghana	MR	Republic of Mauritania	TR	Turkey
BG	Bulgaria	HU	Hungary	MT	Malta	TT	Trinidad and Tobago
BJ	Benin	IS	Iceland	ME	Montenegro	UA	Ukraine
BR	Brazil	IT	Italy	MR	Mauritius	UG	Uganda
BY	Belarus	JP	Japan	NA	Namibia	US	United States of America
CA	Canada	KE	Kenya	NE	Niger	UZ	Uzbekistan
CC	Cocos	KZ	Kazakhstan	NL	Netherlands	VC	Virgin Islands
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	YE	Yemen
CI	Cote d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand	ZW	Zimbabwe
CM	Cameroon	LA	Laos	PE	Peru		
CN	China	LB	Lebanon	PT	Portugal		
CU	Cuba	LC	St. Lucia	RO	Romania		
CY	Cyprus	LI	Liechtenstein	RU	Russian Federation		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

WO 00/65421

PCT/US00/11073

## RECEPTOR SELECTIVITY MAPPING

## BACKGROUND OF THE INVENTION

The present invention relates generally to a combination of chemoinformatics and  
5 bioinformatics and data on chemical-molecular target interactions to create multi-  
dimensional databases. More particularly, this invention relates to databases comprising  
chemical compound, molecular target, and biological or clinical information in which  
patterns or relationships of interactions between chemical compounds and molecular  
targets are determined and compared with other information in the database in order to  
10 draw conclusions that are useful for drug discovery and development and for related areas.

The worldwide pharmaceutical industry spends more than \$30 billion a year on  
research and development, of which nearly one-third is spent on the discovery and early  
development phase, which is the period leading up to the selection of a drug candidate for  
preclinical and clinical development. Some critical steps in drug discovery include (1)  
15 sequencing DNA comprising segments of the human genome; (2) identification of genes  
within the genome that are associated with specific diseases or biological functions; (3)  
production of a protein such as a receptor or enzyme that corresponds to, or is encoded by,  
the functional gene and which then becomes a biological or molecular target for drug  
discovery; (4) screening a library of chemical compounds for activity against the  
20 molecular target (high throughput screening); (5) screening the most potent active  
compounds against other biological targets (particularly other receptors or enzymes) to  
assess the compounds' selectivity or specificity for the intended biological/molecular target  
and potential to cause undesirable side effects through activity at other targets; (6)  
evaluating the most potent and selective compounds for their activity in a range of other  
25 assays designed to measure such properties as toxicity, absorption, distribution,  
metabolism, excretion, etc.; (7) assessing the most promising compounds based on  
empirical judgments using the above information, and then sending that information to a  
chemical synthesis group to produce analogs (or modified but related chemical structures)  
of the initial active compounds; (8) retesting the chemical analogs through Steps (4), (5)  
30 and (6), then repeating Step (7) until an optimized lead compound or series of compounds  
is identified; and (9) forwarding the optimized lead compounds to further preclinical and  
clinical testing.

WO 02/65421

PCT/US00/11073

Throughout this process of discovery and development, compounds go through successively narrower filters, and compounds are eventually selected for the more expensive phases of preclinical and clinical development. Unfortunately, the selection process often leads to preclinical testing and clinical testing of compounds that will fail at these stages and never reach commercialization. These failures lead to extremely high average costs, estimated to exceed \$300 million, to develop and launch a new drug. If, however, the optimal drug candidate is correctly identified early in the discovery and development process and successfully passes preclinical and clinical testing, the actual cost to develop that drug may be reduced by as much as 75%. Clearly, a major goal of pharmaceutical R&D should be to enhance the predictability of early drug development tests such as outlined above.

With the revolution of new techniques in biotechnology and the evolution of tools to automate many laboratory processes, two dominant trends have emerged in recent years that are having an important impact on pharmaceutical R&D. First, the number of molecular targets (such as new receptors and enzymes) available for discovery screening programs continues to increase dramatically due to progress in sequencing the human genome. About 400 molecular targets have been explored for drug discovery; estimates of the number of potential molecular targets that may be elucidated from the human genome project range in the thousands to more than 10,000. Second, the size of chemical compound libraries available for discovery screening programs has expanded nearly tenfold (to more than a million compounds in many drug companies) due to automation and new technologies such as combinatorial chemistry. These two factors hold tremendous promise for new drug discovery, but they also create significant potential problems having adverse consequences on the cost of drug development. More targets and more compounds will result in many more bioactive compounds being discovered, leading to greater difficulty in selecting the optimal drug candidates to advance to preclinical testing, as well as increased development costs due to more compounds entering preclinical and clinical testing and potentially more failures at these stages.

These factors point to an increased need for rapid, inexpensive, *in vitro* ("test-tube" or microplate-based) assays for lead compound selection, optimization, and validation. Such rapid assays may help identify the most promising of these active compounds before



WO 00/65421

PCT/US00/11073

they enter the later, more expensive, stages of drug development. These factors further point to a need for more effective methods to manage and interpret the vast amount of data on genes and gene products (molecular targets), chemical structures, and screening results.

One application of *in vitro* assays that is gaining increased importance in pharmaceutical R&D is "profiling." The assignee of this patent application pioneered the concept of profiling in the late 1980's. Drug companies are provided with an extraordinarily broad array of *in vitro* assays for characterizing the pharmaceutical activity and the potential side effects of compounds under development as new drugs. Currently there are more than 200 different assays that may be performed on a routine basis based on molecular targets, called receptors and enzymes, that play a key role in a wide range of human diseases, including those associated with central nervous system disorders, immune diseases, pain and inflammation, infectious diseases, cancer, metabolism or growth factors, cardiovascular function, and the endocrine system. Pharmaceuticals accounting for more than one-half of the worldwide market function by interacting with cellular receptors. In addition, many side effects of pharmaceuticals are also mediated through their interactions with receptors or enzymes.

Through profiling, a drug company's lead compounds, generally those entering preclinical development, are tested in a battery of receptor and enzyme assays. Information from the profiling process about interactions between the drug company's compound and certain receptors is important for the process of lead compound optimization and selection and can suggest possible side effects or secondary therapeutic activities of the compound. This knowledge can potentially save the drug company millions of dollars in wasted time and expense during preclinical and/or clinical development of the compound.

While profiling services have been practiced for many years, the data generated from these tests are generally used empirically by drug companies. Most drugs, even highly selective drugs, interact with numerous receptors or other molecular targets. Interpreting data produced by profiling, therefore, depends on the experience and knowledge of the scientist from the drug company who reviews the data on both the chemical structure of the compounds and the binding interactions of the compounds with specific receptors. Unfortunately, even the most experienced pharmacologist has an

WO 00/65421

PCT/US00/11073

incomplete knowledge of the interaction of different drug compounds with the broad range of receptors relevant to drug development.

The need for more effective methods to manage, collate, interpret, and utilize the vast amount of data on genes and gene products (molecular targets), chemical structures, and screening results has led to the creation of new opportunities in bioinformatics and chemoinformatics, or managing biological and chemical data. The stages of generating large pools of information for drug discovery can be broken down into (1) DNA sequences (code of genetic material or genes that are blueprints for the cell to make gene products or proteins); (2) functional genomics (process of conversion of DNA sequences to expression of corresponding gene products or proteins via mRNA production, especially in response to drugs or changes in biological function); (3) proteomics (identification of the amino acid sequence and/or three-dimensional structure of gene products or proteins, such as receptors, for which the genes code); (4) small molecule pharmacology/toxicology (molecular binding or interactions between gene products, like receptors, and small organic chemicals that are potential drugs); and (5) chemical structure (of small molecule, drug-like compounds).

Databases for DNA sequences (Group 1) are well established and include GenBank, The Genome Center, and others. Similarly, databases of chemical structures (Group 5) are well known and provided by vendors such as MDL (Isis) and Oxford Molecular. Databases for proteomics (Group 3), such as SWISS-PROT, ProLink, and PDB, are also being established. Each of these databases can be considered as one-component, in that they contain structural information and can be used to determine patterns in that one dimension or single component of structural or sequence information. Databases for Groups 2 and 4 are not well established, but should be valuable additions to the information pool for drug discovery and development. These latter two forms of datasets would be two-component or two-dimensional in that they would contain data relating to the interaction between two structures, such as genes to proteins (Group 2) and proteins to chemicals (Group 4). Such relationship databases add a significant level of complexity compared with the one-component databases.

Partial databases or datasets for Group 4 relationships have been or are being established. For example, profiles of the binding of single compounds against a broad set

WO 00/65421

PCT/US90/11073

of receptor targets by the assignee for its clients is a partial dataset for Group 4-type databases. Similarly, data generated through high throughput screening projects in which thousands to hundreds of thousands of chemicals, such as might be contained in a chemical structure database (Group 5), are screened for activity against a specific receptor target (a single point in a Group 3 database), would represent a partial Group 4 database. Although such partial Group 4 datasets will be helpful aids for drug discovery and development, they suffer from two major drawbacks. First, they are directed toward specific two-component analyses, such as the binding selectivity of a single compound or limited set of compounds across a range of receptors (profile) or of many compounds at one receptor target (high throughput screening). In both cases, the breadth of the dataset is insufficient to allow statistical correlations to be drawn among a multiplicity of receptor targets and a multiplicity of chemical structures. Second, and importantly, these partial datasets are being generated on chemical compounds selected for their structural novelty and therefore proprietary potential as new drugs. Since these are novel compounds, there does not exist any biological information about the activity of these compounds in animals or humans. Such approaches therefore suffer the same limitations as the pharmacologist trying to empirically interpret the data of a profile, as described above.

#### SUMMARY OF THE INVENTION

Accordingly, it is an object of the present invention to meet the foregoing needs by providing systems and methods for analyzing data relevant to drug discovery and development. A full-rank screening database including positive and negative data resulting from a large number of chemical compounds being tested against a large number of molecular targets is provided. The number of combinations of chemical compounds and molecular targets must be large enough such that a person of ordinary skill in the art of statistical or other data mining methods can use the screening database together with the corresponding chemical compound database and molecular target database to produce a reliable prediction of which chemical compounds are suitable for clinical testing and have an enhanced probability to be safe and effective drugs.

Specifically, systems and methods for meeting the foregoing needs are disclosed.

The system includes a computer system comprising a first database containing records corresponding to a plurality of chemical compounds and records corresponding to

WO 00/5421

PCT/US00/11073

biological information related to effects of the plurality of chemical compounds on biological systems of humans or animals, and a second database containing records corresponding to a plurality of molecular targets. The computer system further comprises a third database containing records corresponding to tests of binding, reactivity, or other interactions between compounds in the first database and molecular targets in the second database, the tests including information on the effect that a compound from the plurality of compounds in the first database has on the interaction between a selected compound (e.g., a reference agent or standard) known to interact with a specific molecular target from among the plurality of molecular targets, said tests being performed for a plurality of the molecular targets in the second database. Means for setting an interaction test threshold corresponding to said effect and means for selecting the compound, sets of compounds, and/or information associated with such compound(s) when the results of the testing of the effect meet the interaction test threshold are also included in the computer system. A user interface is provided to allow a user to view and manipulate or analyze information from the first database, the second database, and the third database as it relates to one or more compound records in the first database and/or as it relates to one or more molecular target records in the second database, especially with respect to compounds, molecular targets, or other database records associated with results that meet the interaction test threshold(s).

Furthermore, the invention relates to using methods of statistical analysis and other data mining methods as applied to these multidimensional databases to determine correlations or patterns that are relevant to drug discovery and development.

Both the foregoing general description and the following detailed description provide examples and explanations only. They do not restrict the claimed invention.

#### DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate embodiments of the invention and, together with the description, explain the advantages and principles of the invention.

Fig. 1A illustrates a chemical compound table in the receptor selectivity mapping database according to one embodiment of the present invention;

WO 03/65421

PCT/US00/11073

Fig. 1B illustrates a snap-shot of a chemical compound record containing spatial coordinates of a compound in the receptor selectivity mapping database according to one embodiment of the present invention;

Fig. 2 illustrates several logical tables that may be used to access the molecular target information in the receptor selectivity mapping database according to one embodiment of the present invention;

Fig. 3 illustrates a biological information table in the receptor selectivity mapping database according to one embodiment of the present invention;

Fig. 4 illustrates the use of a receptor selectivity mapping database as part of a screening process according to one embodiment of the present invention;

Fig. 5A illustrates the use of a receptor selectivity mapping database as part of a screening process to discover and select new compounds as potential new drug candidates for further development;

Fig. 5B illustrates the use of a receptor selectivity database as part of a screening process to identify new targets as potential validated targets to use to discover new drug candidates for specific disease indications;

Fig. 6A illustrates the use of a database for predicting the drug potential of a new compound; and

Fig. 6B illustrates the use of a database to validate the disease relevance and/or the biological function of a new molecular target.

#### DETAILED DESCRIPTION

Reference will now be made to preferred embodiments of this invention, examples of which are shown in the accompanying drawings and will be obvious from the description of the invention. In the drawings, the same reference numbers represent the same or similar elements in the different drawings whenever possible.

Systems and methods consistent with the present invention allow the analysis of data relevant to drug discovery and development, for example, for predicting the potential of a new compound's suitability for progression to preclinical and clinical tests with an enhanced probability of becoming a safe or effective new drug. For purposes of the following description, the systems and methods consistent with the present invention are described with respect to a relational database containing multiple main tables and with the

WO 00/65421

PCT/US00/11073

use of the binding between chemical compounds and molecular targets as a measurement of the interactions between the two. The description should also be understood to apply in general for any database structure having multiple main components and to the measurement of any interactions between chemical compounds and molecular targets.

- 5 The present invention relates to the novel design, construction, and application of a database relating information-rich chemicals, molecular targets, especially proteins or other macromolecules, and biological activity of the chemicals. Furthermore, the present invention relates to the primary use of known drugs and drug candidates that have failed in clinical or preclinical trials as a source of the chemical library for the database, together  
10 with preclinical or clinical data generated for such chemicals describing their side effects, mechanism of action and other medically relevant data. The present invention further relates to determining the binding or other interactions between the chemicals and the molecular targets in the database, then using methods of relationship analysis and data mining to correlate patterns of these interactions with specific biological activities that are  
15 relevant to drug discovery and development, or with specific chemical structures, substructures, or other features of compounds exhibiting such interactions, or with biochemical, structural, or other features of molecular targets exhibiting such interactions. Examples of such data mining techniques can be found in the following references, which are incorporated by reference in their entirety:
- 20 a) Chen *et al.*, Recursive Partitioning Analysis of a Large Structure-Activity Data Set Using Three-Dimensional Descriptors, *Journal of Chemical Information and Computer Sciences*, October 1998;
- b) Hawkins *et al.*, Analysis of a Large Structure-Activity Data Set Using Recursive Partitioning, *Quant. Struct.-Act. Relat.*, 16:296-302 (1997);
- 25 c) DePriest *et al.*, 3D-QSAR of angiotensin-converting enzyme and thermolysin inhibitors; a comparison of CoMFA models based on deduced and experimentally determined active site geometries, *J. Am. Chem. Soc.*, 115:5372-84 (1993);
- d) Good *et al.*, in *Reviews in Computational Chemistry*, Lipkowitz, K. B., Boyd, D. B. (eds.), VCH, New York, Vol. 7, pp 67-117 (1996);
- 30 e) Marshall *et al.*, in *Computer-Assessed Drug Design*, ACS Symposium Series 112, American Chemical Society: Washington, DC, 1979, pp 205-226;

WO 00/65421

PCT/US00/11073

- f) Moloc *et al.*, A three-dimensional structure activity relationships and biological receptor mapping, in *Mathematics and Computational Concepts in Chemistry*, Ellis Horwood, Chichester, 1985; pp 225-231;
- g) Mayer *et al.*, A unique geometry of the active site of angiotensin-converting-enzyme consistent with structure activity studies. *J. Comput. Aided Mol. Des.*, 1:3-16. (1987);
- h) Sheridan *et al.*, The ensemble approach to distance geometry: application to the nicotinic pharmacophore, *J. Med. Chem.* 29:899-906 (1986);
- i) Martin *et al.*, A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists, *J. Comput. Aided Mol. Des.*, 7:83-102 (1993);
- j) Catalyst/Hypo Tutorial, version 2.0, BioCAD Corp. Mountain View, CA, 1993
- k) Sprague, P. W., Automated chemical hypothesis generation and database searching with Catalyst. *Perspect. Drug Discov. Des.*, 3:1-20 (1995);
- l) Burnum *et al.* Identification of common functional configurations among molecules, *J. Chem. Inf. Comput. Sci.*, 1996, 36:563-71 (1996).
- m) HipHop Tutorial, version 2.3; Molecular Simulation Inc.; Sunnyvale, CA, 1995;
- n) Davies, K. and Upinn, R., 3D pharmacophore searching, *Net. Sci.* (<http://www.netsci.org/Science/Cheminform/feature02.html>);
- o) Golender, V. and Vesterman, B., APEX 3D expert system for drug design, *Net. Sci.* (<http://www.arwod.com/netsci/Science/Compchem/feature09.html>);
- p) Van Drie, J., Strategies for the determination of pharmacophoric 3D database queries, *J. Comput. Aided Mol. Des.*, 11:39-52 (1997);
- q) Van Drie, J. and Nugent, R., Addressing the challenges posed by combination chemistry: 3D databases, pharmacophore; recognition and beyond. *SAR QSAR Environ. Res.*, 9:1-21 (1998);
- r) Finn *et al.*, Pharmacophore discovery using the inductive logic programming prolog, in *Machine Learning. Special Issue on Applications and Knowledge Discovery*, Kluwer Academic Publishers: Boston, 1998, pp 1-33; and
- s) Jain *et al.*, Compass: a shape-based machine learning tool for drug design. *J. Comput. Aided Mol. Des.*, 8:635-52 (1994).

WO 00/65431

PCT/US00/11073

The background section suggests that, contrary to standard operating procedures in the pharmaceutical industry, a Group 4 database should be established having more components than a two-component database, and that it should cover a substantial breadth of both receptor or enzyme targets and chemical compounds. By way of example, a three-component database would be created by first selecting a broad set of chemical compounds that are rich in information of direct relevance to drug discovery and development. The most relevant information is often obtained by actual experience of testing such chemical compounds in humans through clinical trials and/or post-marketing surveillance or in animals through preclinical testing. Other relevant biological information may come from natural products that demonstrate one or more observed bioactivities, as well as chemical reference standards that have been used in the industry to characterize the biology of receptors. Accordingly, one embodiment of information-rich chemical compounds selected for such a Group 4 database includes marketed pharmaceuticals, drugs that have failed in clinical or preclinical trials, bioactive natural products or natural extracts, and reference agents used for receptor binding assays.

One may construct such a database using screening data obtained from the scientific literature. While this approach could yield partial datasets, it may have limitations. First, literature references generally provide only positive information (e.g., reports of inhibition of binding of a specific compound to a specific receptor) and not negative data (e.g., a lack of inhibition of binding and therefore lack of activity). In determining useful comparisons of information, negative data can be as valuable as positive data. Furthermore, certain statistical analyses may not be applicable to datasets that lack completeness of both positive and negative data. Second, separate quantitative reports of binding data for one compound against a receptor in one article vs. reports of binding data for a second compound at the same receptor may not be comparable because of variations in the way the assays were performed. Therefore, one embodiment for creation of a Group 4 three-component database would be to screen a broad array of compounds through a broad array of receptor or enzyme targets in order to obtain consistent comparative results and ensure the collection of both positive and negative data.

#### The Chemical Compound Component: Selection of Chemical



WO 00/65421

PCT/US00/11073

**Libraries and Inclusion of Chemical Data**

The present invention relates to databases that contain, as one component, chemical compounds about which information is known concerning biological activity relevant to pharmaceutical research and development. The biological activity information may be included in the chemical compound database or table.

For example, these information-rich chemicals include:

(a) Compounds that are pharmacological reference agents or reference standards for measuring the interaction or molecular binding between unknown chemical compounds and a specific molecular target, such as a receptor or enzyme. Examples of such reference compounds include those compounds that are used for characterizing binding interactions between test compounds and molecular targets including receptors or enzymes. Other reference agents could include chemicals selected from the catalog of Research Biochemicals Inc. (RBI), a unit of Sigma Aldrich Corp., and from other sources that are well known in the industry. These pharmacological reference compounds often have been tested previously and/or marketed as pharmaceuticals or are natural products with characterized biological activity and therefore may overlap with compounds in the following three categories;

(b) Compounds that are known pharmaceuticals that are currently or have previously been marketed for clinical use, and for which there is a substantial amount of biological information available. These compounds are well-known and are listed in publications available from U.S. government agencies such as the Food and Drug Administration (FDA), as well as publications by private or non-profit organizations. One such publication by a non-profit organization is the United States Pharmacopeial Convention Inc.'s *USP DI Series*, including *Volume I. Drug Information for the Health Care Professional*, which is updated monthly by *USP DI Update*. As new drugs are approved for marketing, they would be included in this category. Marketed pharmaceuticals or drugs approved by the FDA or equivalent foreign regulatory bodies are a matter of public record so that one normally skilled in the art can easily identify chemical compounds that would be included in this category;

(c) Compounds that have been approved for testing in humans, such as compounds that had been granted IND (Investigational New Drug) status, as potential drugs but that

WO 00/65421

PCT/US00/11073

failed to achieve sufficient efficacy or safety in clinical trials to gain approval from the FDA or otherwise did not reach the status of marketed pharmaceuticals. Compounds in this category may also include those compounds that have been approved by the FDA for commercialization but that have later been withdrawn from the market. These compounds  
5 also would have a significant amount of biological information available and would be especially useful for purposes of this invention. The identity of failed drugs can be obtained from numerous sources, including public announcements by drug and biotechnology companies, publications such as the "Pink Sheets," and lists maintained by the FDA; and

10 (d) Compounds that are obtained from natural sources such as plants, microorganisms, animals, etc., that exhibit biological activity. These natural products may include toxins, antimicrobial agents, behavioral modifiers, defensive agents, and other categories of compounds that provide information relevant to drug discovery and development. The identity of natural products can be found in numerous publications,  
15 including but not limited to, the RBI catalog and Sigma Aldrich catalog of chemical compounds.

For each compound included in the database, chemical structure, chemical formulae, physical-chemical characteristics, chemical space coordinates or other chemical structure descriptors (e.g., Smiles codes), solubility, and other relevant data, to the extent  
20 such information is available, are entered into fields in the database. Those skilled in the art would recognize other parameters that might be included. Chemicals can be organized by chemical structure relatedness in the database or by other relationships.

Fig. 1A illustrates a chemical compound table 300 in a relational database system. The table 300 lists a number of chemical compounds and includes records (rows 1-N) of a  
25 number of compounds N. For each compound there may be a number of corresponding columns 301-307 containing information related to the compound. For example, in Fig. 1A column 301 contains the name of the compound; column 302 includes the compound type (e.g., compounds that have been approved for testing in humans, etc.); column 303 includes information related to the chemical structure, for example, a hyperlink that brings  
30 up a screen containing a drawing of the structure (see snap-shot 310 in Fig. 1B); column 304 includes the chemical formula for the compound; column 305 includes information

WO 00/55421

PCT/US00/11073

about the physical-chemical characteristics of the compound; column 306 includes chemical space coordinates of the compound; and column 307 includes solubility information of the compound.

Additional columns may be added in order to include other relevant data related to each chemical compound 301 listed in the table 300. These additional columns may include biological activity of the compound, rendering the chemical compound database a two component database (see also database 500).

Fig. 1B illustrates a snapshot 310 that may include information corresponding to a record in the table 300. For example, the chemical formula 304 of a compound may be included in the snapshot of the record as well as the compound's structure 303.

**The Molecular Target Component: Selection of Receptors, Enzymes, and Other Molecular Targets and Inclusion of Molecular Target Data**

Molecular targets such as receptors, enzymes, other proteins, nucleic acids, carbohydrates, and other macromolecules relevant to drug discovery and development, are representative of the second component of the databases comprising this invention. In one embodiment of this invention, receptors and enzymes are the principal molecular targets. Receptors mediate much of the molecular communication among cells and organs in the body. Enzymes often amplify such communications through, for example, secondary messenger systems and cell signaling pathways.

Receptors include classical families of receptors such as dopamine receptors, serotonin receptors, opiate receptors, muscarinic receptors, adrenergic receptors, adenosine receptors, etc. These receptor groups include subtypes of the receptor type (such as dopamine-1, dopamine-2, dopamine-3, dopamine-4, and dopamine-5 receptors). Certain subtypes have further variations (such as dopamine 4.2, dopamine 4.4, and dopamine 4.7) or can have different forms (such as dopamine 2 short and dopamine 2 long). Splice variants of receptors can also occur, as can mutations in the genes encoding specific receptors which might lead to a subset of a population that has a receptor with slightly different binding affinity for drugs or other compounds compared with the normal receptor type. Receptors can be grouped by family, superfamily, or subfamily. Some groupings include G-Protein Coupled Receptors, 7 transmembrane receptors, nuclear receptors, etc.

WO 00/65421

PCT/US00/11073

Receptors can be grouped by the degree of homology of the DNA sequence of their corresponding genes. Receptors can also be grouped by their amino acid sequence and related three-dimensional conformations. Receptors can be classified by their location of expression in tissues or across different cell types.

5 Enzymes can include proteases, carbohydrases, kinases, phosphatases, DNA-modifying enzymes, transferases, P450's, and others known to those skilled in the art.

Other receptors, receptor sources, and corresponding assays are constantly being developed by the assignee to be added to the content of the database. Additional receptors and receptor assays are well known to those skilled in the art. Lists and descriptions of  
10 certain receptors relevant to drug discovery and development can be found in numerous publications known to those skilled in the art. These publications include the RBJ Handbook of Receptor Classification and the IUPHAR receptor classification book. Furthermore, as new receptors and receptor subtypes are discovered, they can be added to the content of the database.

15 Enzymes and enzyme assays are well known to those skilled in the art. Lists and descriptions of certain receptors relevant to drug discovery and development can be found in numerous publications known to those skilled in the art.

Fig. 2 illustrates tables 400, 410, and 420 forming part of a relational database system which may be used to access molecular target information. Table 400 lists the  
20 targets and includes records (rows I-M) of a number of targets M. Column 401 lists the names of the target, while column 402 specifies the target type corresponding to each target name.

Table structures may vary according to the target type specified in column 402. Table 410 includes information about those targets listed in table 400 which are classified  
25 as receptors. Records from table 410 may be accessed by querying the database for a particular receptor name. The receptor names found in table 410 may be accessed, in turn, by querying table 400 for those target names for which column 402 reads "Receptor."

In table 410, column 411 contains the name of the receptor, which is also the name of the target in column 401 in table 400; column 412 includes receptor family information;  
30 column 413 includes receptor superfamily information; column 414 includes receptor subfamily information; column 415 includes the information about the degree of

WO 00/65421

PCT/US00/11073

homology of the DNA sequence of corresponding genes; and column 416 includes information on amino acid sequence. The amino acid sequence is one of a number of molecular descriptors that may be included in the database. Other molecular descriptors 417, for example, could include hydrophathy plots corresponding to the amino acid  
5 sequence. Because the molecular target database represented by tables 400, 410, and 420 includes target information and associated biological information related to the targets is included in the database (see table 600), this database may be considered a two-component database. The columns shown are illustrative of the types of information that may be included in the database and should not be construed as limiting the invention.

10 Table 420 includes information about those targets in table 400 that are classified as enzymes. Records from table 420 may be accessed by querying the database for a particular enzyme name. The enzyme names found in table 420 may be accessed, in turn, by querying table 400 for those target names for which the target type column 402 reads "Enzyme."

15 In table 420, column 421 contains the name of the enzyme, which is also the name of the target in column 401 of table 400 and column 422 includes enzyme type information. Column 423 is labeled as "Other relevant information" and is included in the table for purposes of illustrating that additional columns may be added to table 420 depending on other enzyme information that a user of the database might want to access,  
20 including amino acid sequence and molecular descriptors.

Although only tables 410 and 420 are shown to describe the access of molecular target information by using the target type, additional tables may be added to the relational database system corresponding to the number of molecular target types available in the database.

25

**The Biological Information Component: Selection of  
Biological/Clinical Information Parameters**

Biological information forming part of the database includes material that would  
30 relate to side effects, mechanism of drug action, metabolism of a drug, toxicity, adsorption, distribution, and excretion, for example. This information is available on FDA-approved labels of marketed drugs, or from literature sources and publications for drugs that have

WO 00/65421

PCT/US00/11073

- failed in clinical trials. Examples of some specific parameters are toxicity, LD<sub>50</sub>, LD<sub>01</sub>/ED<sub>50</sub>, teratogenicity, mechanism of toxicity, target organ for toxicity, in vitro toxicity battery, induction of apoptosis, bioavailability, absorption, blood-brain barrier, oral absorption, mucosal absorption, % absorbed, distribution, blood protein bound, half-life,
- 5 onset of action, duration of action, peak concentration in blood, metabolism, major pathway, minor pathway, active metabolites, excretion, primary excretion mode, secondary excretion modes, *in vivo* effects, therapeutic indication, animal behavioral effects, side effects, primary known target, other organ/system targets, and known receptor interactions.
- 10 Fig. 3 shows table 500 which includes some of the biological information parameters mentioned above. Table 500 comprises N rows (1 through N) which correspond to all the possible chemical compounds in the first database. Column 501 includes the compound name; column 502 includes the therapeutic indication (for marketed or failed drugs); column 503 includes toxicity information; column 504 includes
- 15 side effects information; and column 505 includes information on the mechanism of drug action. Table 500 would be associated with table 300, for example, to form a two-component chemical compound and biological activity table.
- Fig. 3 also shows table 600, which includes biological information parameters associated with the molecular targets in the database. Table 600 includes P rows (1
- 20 through P) which correspond to all the possible targets in the second database. Column 601 includes the target name; column 602 includes the therapeutic indication (for marketed or failed drugs); column 603 includes toxicity information; and column 604 includes side effects information. Similarly, table 600 would be associated with table 400, for example, to form a two-component molecular target and biological activity table. Tables 500 and
- 25 600 together may be a full-rank database (*e.g.*, including all possible combinations between compounds and molecular targets in a relational database system) including molecular target information, chemical compound information, and biological activity information associated with each of the molecular targets and with each of the chemical compounds, and may be considered a multidimensional database. Additional columns
- 30 may be included in tables 500 and 600 without departing from the invention.

WO 00/65421

PCT/US00/11073

Determining Binding Information

A key feature of this invention is the establishment of several components of information which, by way of illustration, comprise chemicals, molecular targets, and biological information, and measuring the binding, reactivity or other interactions between the chemicals and molecular targets. This binding or reactivity information can then be related back to the known biological information in order to distinguish patterns and relationships that can be used for drug discovery and development. An important aspect of this invention is to generate broad and consistent binding or reactivity data between the chemicals and molecular targets in order to provide as complete a dataset as possible in order to be able to identify relevant patterns or relationships and to provide both positive and negative binding or reactivity information for the datasets. In one embodiment, the binding data is established as a numerical descriptor that either satisfies or does not satisfy a threshold set, for example, for a specific molecular target or set of molecular targets. The numerical descriptor may relate to the activity or lack of activity for each compound and each receptor or other molecular target measured at a concentration deemed near the appropriate threshold for relevance to the biological system or biological information set. For example, chemicals can be tested at  $10^{-5}$  M (10 micromolar) for their ability to inhibit binding at a threshold of 30% between a receptor and its specific reference compound. Other initial concentrations or percentage inhibition thresholds can be selected. Also, in one embodiment, those chemicals that demonstrate inhibition of binding above the threshold in the initial yes/no testing are further tested for the potency of the binding inhibition. These active chemicals are tested at a series of concentrations that might, for example, include tests at 7-14 different concentrations within the range of  $10^{-5}$  to  $10^{-9}$  M, such that an  $IC_{50}$  and/or  $K_i$  value can be determined for the active compound at the specific receptor. Fewer or more concentrations may be used for such determinations and concentrations above or below  $10^{-5}$  to  $10^{-9}$  M may be required. These data then yield a matrix of relative degree of activity or relative potency for each active compound at each molecular target.

In order to generate these screening data, chemicals are first solubilized in a suitable solvent system, such as 4% DMSO, although other concentrations of DMSO and other solvents are also acceptable. These chemical stock solutions are then diluted to the

WO 00/55421

PCT/US00/1073

appropriate concentration and made available as repositories. For each assay measuring the interactions between the chemical and molecular target, the reagents and protocols for the assay will vary. Each such assay needs to be characterized and routinely established for consistency. Appropriate controls need to be run each time the assay is performed.

- 5 Any assay format that can generate the desired type and accuracy of information can be used. Numerous assay detection systems, such as radioactive labels, fluorescence, fluorescence polarization, time-resolved fluorescence, fluorescence correlation spectroscopy, chemiluminescence, UV absorption, colorimetric, *etc.*, can be used.

- In one embodiment, a receptor-binding assay or enzyme activity assay is used to  
10 generate data on molecular interactions. As an example, for a receptor binding assay, chemicals from a repository are tested for their ability to inhibit the binding interaction between the receptor and a reference agent selected for that receptor. The receptor may be derived from a tissue source, such as animal or human tissue, or from a cell line expressing the receptor, or from a transfected cell line containing the gene for the receptor. The  
15 receptor source is prepared for the assays, for example, by preparing a membrane fraction containing the receptor. Alternatively, the receptor may be partially purified. The reference compound, or ligand, is preferably selected for its potent and/or specific binding to the specific receptor and may have a radioactive tracer such as Iodine-125 or tritium or carbon-14 or other marker to enable a bound ligand to be distinguished from an unbound  
20 ligand. Coincident with testing the chemicals for binding data to include in the database, positive and negative controls are run, as is a reference curve with varying concentrations of the reference (radio)ligand to ensure the quality of the assay run.

- A plurality of methods and systems may measure the interactions between targets and compounds as would be recognized by a person of ordinary skill. The radioligand,  
25 receptor preparation, and test compounds are incubated together for an appropriate time, in an appropriate buffer, and at an appropriate temperature, often with the objective of reaching equilibrium of the binding reactions. The amount of bound versus unbound radioligand is determined by a separation step, such as filtration, or by use of a method, such as SPA (scintillation proximity assay), and measured by liquid scintillation or gamma  
30 counting. The amount of specific binding of the test compound is then determined by



WO 00/65421

PCT/US00/11073

comparing assay results for the test chemical(s) vs. the positive and negative controls. The per cent inhibition of the test chemical(s) is calculated from these data.

Fig. 4 shows table 200 as an illustration of a screening results and assay database in which, for example, chemical compounds included in database 300 (comprising 1 to N chemical compounds) are tested for their effect against molecular targets included in database 400. Numerous forms of table 200 are possible. For example, in table 210 screening results are entered in a "yes" or "no" entry with respect to whether the screening result for each of a plurality of chemical compounds tested against each of a plurality of molecular targets was above or below the selected threshold test result for each set of determinations.

As another example, in table 220 screening results are entered as a numerical descriptor identifying the potency or magnitude of the binding or other effect (e.g., the  $K_i$  for chemical:receptor interactions) for each of a plurality of chemical compounds tested against each of a plurality of molecular targets. In a preferred embodiment, all such matrix points for chemicals x targets in tables 210 and 220 are determined and entered into the database such that a full-rank dataset is derived. The screening results and assay database 200 may also include other measurements of chemical:target interactions, including raw data of screening results and measurements derived from the raw data, assay protocols and performance characteristics, and other relevant information.

Figs. 5A and 5B illustrate the use of a database 100, here shown as a receptor selectivity database, by way of example, as part of a screening process to discover and select new compounds as potential new drug candidates for further development (Fig. 5A) or new targets as potential validated targets to use to discover new drug candidates for specific disease indications (Fig. 5B). The database 100 may include a chemical compound component 300; a molecular target component 400; biological information components 500 and 600; and a screening results and assay database 200.

A new compound or set of compounds is introduced to a screening process 102 for determining whether it is effective in inhibiting the binding of a specific chemical compound (e.g., a reference agent) and a molecular target (see Fig. 5A). The screening process may use target information from the molecular target component 400.

WO 00/55431

PCT/US00/11073

The results of the screening process 102 may be stored in an intermediate database or entered into the screening results and assay database 200 of the receptor selectivity database 100. The results may also be stored in the biological information database 500 as particular parameters (e.g., cytotoxicity, etc.) as well as in the chemical compound database 300 (e.g., name of the compound, etc.).

The complete set of results from the screening process 102 may be stored in the screening results and assay database 200. The database 200 may be queried for those new compounds that exhibit an inhibitory effect on the binding of molecular targets and chemical compounds (e.g., reference agent) so that those new compounds can further be tested.

Alternatively, a new molecular target, such as, for example, an "orphan" receptor about which the structure is known but the function or disease relevance is not known, is introduced to a screening process to be tested against the chemical compounds in the chemical compound database 300 (see Fig. 5B). Results of the screening process, including identification of chemicals that interacted with the new molecular target, are incorporated into the screening results database 200. Queries are made within database 100 to determine further steps to identify the function of the new molecular target and/or validate the disease relevance of the new target.

Fig. 6A illustrates the use of the database 100 for predicting the drug potential of a new compound. A table 710 relies on information from the chemical compound (300), molecular target (400), biological information (500 and 600), and screening results (200) databases. The table 710 is filled in with information from one or more of these databases (or tables) by executing an automatic query script to retrieve the information once a user provides the database 100 with information about a new chemical compound.

The query script used for the creation of table 710 may select chemical compounds from the chemical compound database 300 upon receiving the new compound information. The selection may be based on similar characteristics, such as chemical structure or other properties, between the new compound and the compounds already included in the database 300.

After the selection of chemical compounds, the query script selects targets from the target database 400 that are known to react (e.g., bind) with the selected compounds.

WO 00/65421

PCT/US99/11073

Finally, the combination of selected chemical compounds and selected molecular targets may be used for querying the biological information databases 500 and 600 and inserting biological information corresponding to chemical compound-molecular target pairings into table 710. Alternatively, the user may enter a specific biological information category of interest (e.g., toxicity) so that the biological information included in table 710 is limited to that category.

The table 710 may be queried by the user to produce information relevant to the predictability of the potential use of the new compound as a drug. An example of this would be a query of the molecular targets known to react with chemical compounds associated with the new compound, and the known side effects produced by the chemical compounds when combined with the retrieved targets.

Fig. 6B illustrates the use of the database 100 to validate the disease relevance and/or the biological function of a new molecular target using an approach similar to that used to predict the drug potential of a new compound, but with the data inputs and queries shown in Fig. 6B.

All patent, patent applications, and publications mentioned are incorporated by reference in their entirety into this application.

The foregoing description of embodiments of the present invention provides an exemplary illustration and description, but is not intended to be exhaustive or to limit the invention to the precise form disclosed. Modifications and variations are possible in light of the above teachings or may be acquired from practice of the invention.

WO 00/65421

PCT/US00/11073

**WHAT IS CLAIMED IS:**

1. A computer system comprising:
  - a first database containing records corresponding to a plurality of chemical compounds and records corresponding to biological information related to effects of such chemical compounds on biological systems;
  - a second database containing records corresponding to a plurality of molecular targets;
  - a third database containing records corresponding to tests of interactions between compounds in the first database and molecular targets in the second database, the tests including information on the effect that a compound from the plurality of compounds has on the interaction of a compound known to interact with a molecular target from the plurality of molecular targets and said molecular target; and
  - a user interface allowing a user to view the selected compound and to selectively view information from the first database, the second database, and the third database as it relates to a compound record in the first database or as it relates to a molecular target in the second database.
2. The computer system of claim 1, wherein the interaction includes binding and the effect includes inhibitory effect.
3. The computer system of claim 1, wherein the chemical compounds include compounds with no known biological activity or that have failed in tests.
4. The computer system of claim 1, wherein the chemical compounds include compounds tested in animals.
5. The computer system of claim 1, wherein the chemical compounds include compounds known to have an effect on the environment.
6. The computer system of claim 1, wherein the chemical compounds include pharmacological reference agents.
7. The computer system of claim 1, wherein the chemical compounds include known pharmaceuticals in the market for clinical use for which there is a substantial amount of biological information available.
8. The computer system of claim 1, wherein the chemical compounds include compounds approved for testing in humans.

WO 00/65421

PCT/US00/11073

9. The computer system of claim 1, wherein the chemical compounds include compounds obtained from natural resources that exhibit biological activity.
10. The computer system of claim 1, wherein the molecular targets include receptors.
- 5 11. The computer system of claim 1, wherein the molecular targets include enzymes.
12. The computer system of claim 1, wherein the molecular targets include nucleic acids.
13. The computer system of claim 1, wherein the molecular targets include carbohydrates.
- 10 14. The computer system of claim 1, wherein the records of the first database corresponding to a plurality of chemical compounds are organized in categories related to the description and properties of the compounds.
15. The computer system of claim 14, wherein the categories include:
- 15 compound name;  
compound type;  
physical-chemical characteristics;  
chemical space coordinates or structural descriptors; and  
solubility.
- 20 16. The computer system of claim 1, wherein the first database includes a natural product database.
17. The computer system of claim 1, wherein the first database includes a failed drug database.
18. The computer system of claim 1, wherein the first database includes a chemical registry database.
- 25 19. The computer system of claim 1, wherein the second database includes a three-dimensional structure database.
20. The computer system of claim 1, wherein the second database includes a sequence/mutation database.
- 30 21. The computer system of claim 1, wherein the second database includes a genomic database.

WO 00/65421

PCT/US00/11073

22. The computer system of claim 1, wherein the records in the third database corresponding to biological information related to the chemical compounds effects on the biological targets, are organized in categories that include:

- compound name;
- 5 target name;
- toxicity;
- side effects; and
- mechanism of drug action.

23. The computer system of claim 1 further comprising means for setting an interaction test threshold corresponding to said effect and means for selecting the compound when its use results in a test meeting the interaction test threshold.

24. A method for analyzing data relevant to drug discovery and development comprising:

- 15 selecting chemical compounds from a first database containing records corresponding to a plurality of chemical compounds;
- selecting molecular targets from a second database containing records corresponding to a plurality of molecular targets;
- producing information corresponding to the interactions between each of the selected chemical compounds and each of the selected molecular targets;
- 20 selecting a biological activity from a third database containing records corresponding to biological information related to effects of chemical compounds on biological targets; and
- using the produced information to correlate patterns of interactions between chemical compounds and molecular targets associated with the selected biological activity.

25

25. The method of claim 24, wherein the step of producing information includes the steps of:

- generating binding data of the binding between each of the selected chemical compounds and each of the selected molecular targets by monitoring the inhibitory effect
- 30 that an unknown compound has on said binding;
- setting a binding test threshold corresponding to the inhibitory effect; and

WO 00/65421

PCT/US00/11073

generating information on the combination of unknown compound, molecular target, and chemical compound that meets or fails to meet the binding test threshold.

26. The method of claim 25, wherein the binding data comprises positive and negative binding information.

301		302	303	304	305	306	307
Compound Name	Type	Chemical Structure	Chem Formulae	Physical-chemical characteristics	Chemical space coordinates	Solubility	
1							
2							
3							
4							
5							
6							
	.	.	.	.	.	.	.
	.	.	.	.	.	.	.
N							

FIG. 1A



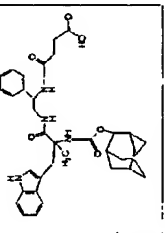
Chemical Snap-shot -		General Description	
Name & Trade Names		Known Target(s) & Activities	
		Receptor Selectivity Mapping	
Formula: $C_{14}H_{14}N_2O_4$		Pharmacology & Toxicology	
M.W.		CAS	
303		Literature	
304		Patent(s)	
Other physical parameters:		Manufacturer:	

FIG. 1B

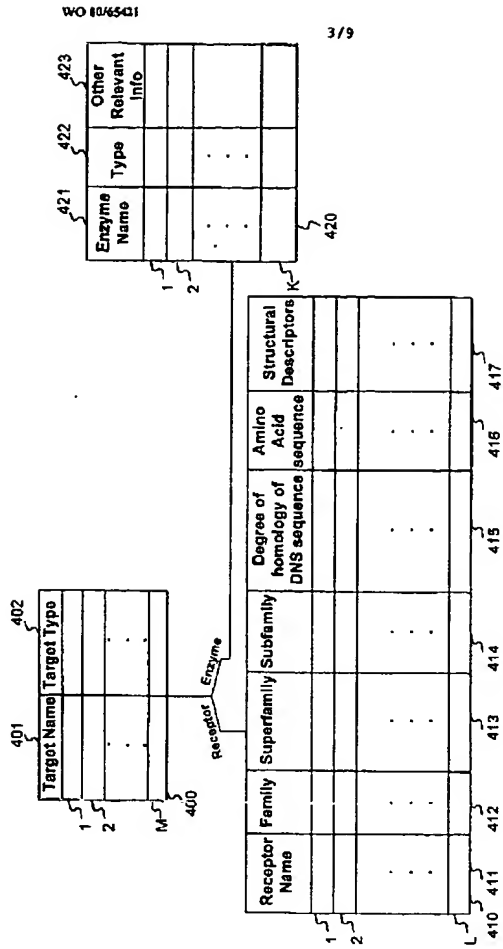


FIG. 2

WO 00/65421

PCT/US98/11073

4/9

Biological/Clinical Information Tables

501		502		503	504	505
Compound Name	Therapeutic Indication	Toxicity	Side Effects	Mechanism of drug action	601	
1	Comp 1				Target Name	603
2	Comp 2				Target 1	604
N					Target 2	
	Comp N				Target P	

FIG. 3

WO 00/65421

PCT/US00/11073

5/9

Compound Name	Target Name	First Pass	Second Pass	IC <sub>50</sub> /K <sub>i</sub>	Threshold	Assay Condition-1
Comp 1						
Comp 2						
Comp N						

200

Target Name Compound Name	Target 1	Target 2	Target P
Comp 1	X	X	X
Comp 2	X	X	X
Comp N	X	X	X

210 X= screening result (above or below threshold)

Target Name Compound Name	Target 1	Target 2	Target P
Comp 1	Y	Y	Y
Comp 2	Y	Y	Y
Comp N	Y	Y	Y

220 Y= screening result (potency e.g. K<sub>i</sub>)

FIG. 4

WO 00/5421

PCT/US00/11073

6/9

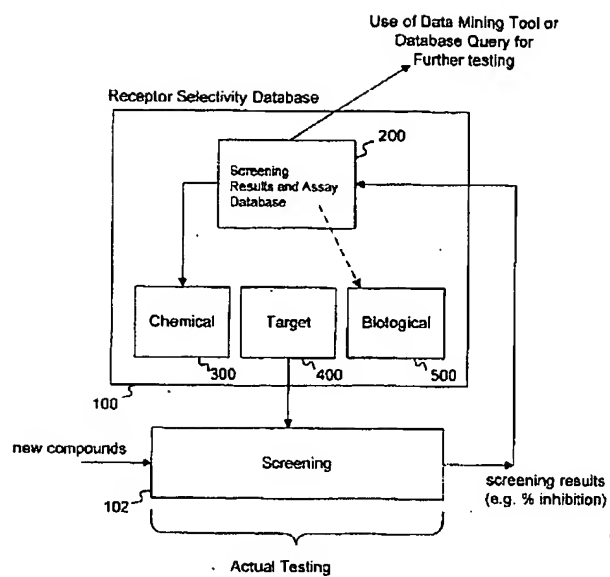


FIG. 5A

WO 00/5421

PCT/US02/11073

7/9

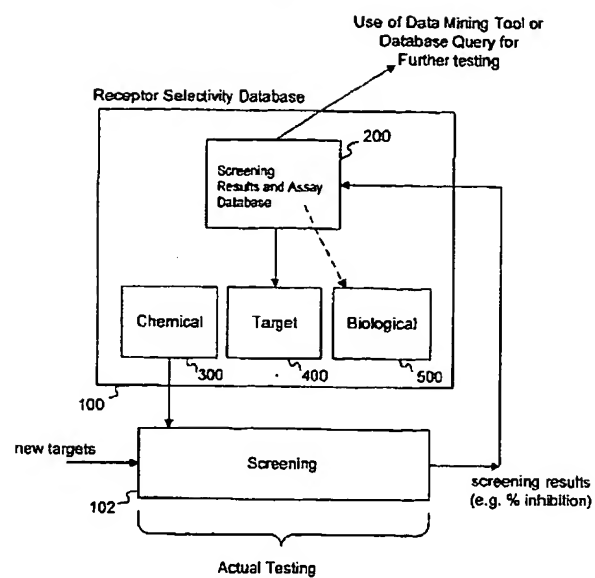


FIG. 5B

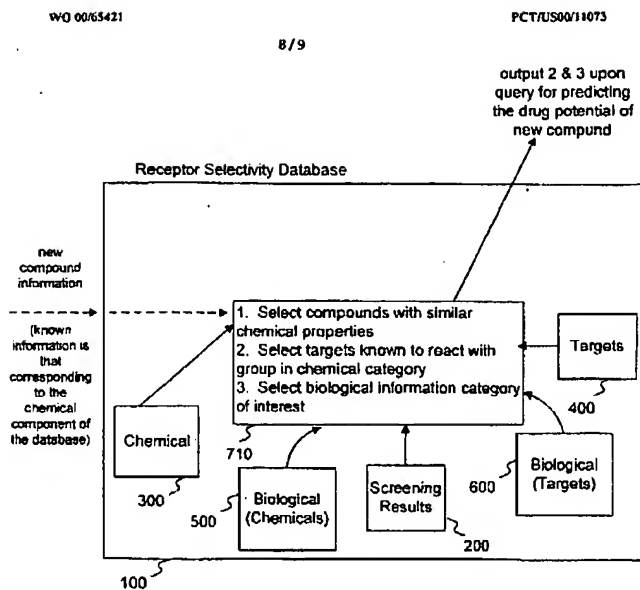


FIG. 6A

WO 00/65421

9/9

PCT/US00/11073

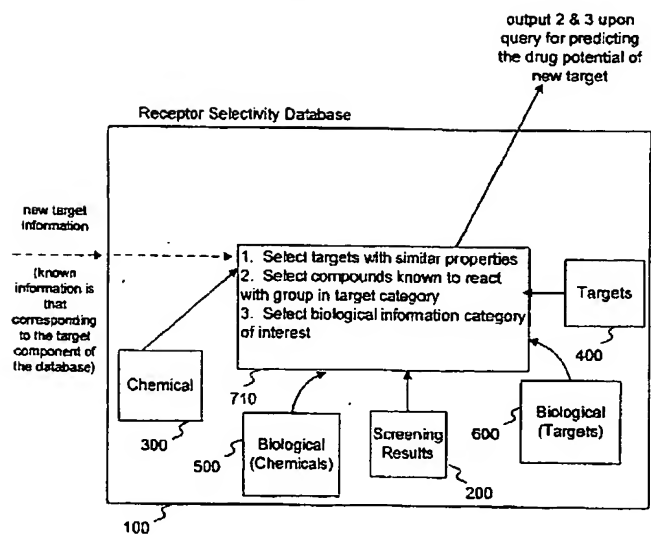


FIG. 6B



## フロントページの続き

(51)Int. Cl.<sup>7</sup>

F I

テーマコード (参考)

// C 1 2 N 15/09

C 1 2 N 15/00

A

(81)指定国 AP(GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), EP(AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW

(72)発明者 マイヤク, デービッド, エム.

アメリカ合衆国 21042 メリーランド州, エリコット シティー, フォリー クアーター  
ロード 12601

(72)発明者 ゼッペテロ, レニー, エー.

アメリカ合衆国 21210 メリーランド州, バルティモア, リッジメデ ロード 310

(72)発明者 チェン, ハオ

アメリカ合衆国 21044 メリーランド州, コロンビア, オスロコート 5905

(72)発明者 ウェイッスマン, アーサー, ディー.

アメリカ合衆国 21215 メリーランド州, バルティモア, メンロ ドライブ 3706

(72)発明者 ラング, ギャリー, エル.

アメリカ合衆国 21014 メリーランド州, ベル エアー, サザン ドライブ 910

F ターム(参考) 2G045 AA29 AA40 CB17 DA12 DA13 DA14 DA20 DA30 JA01

4B024 AA11 CA04 CA06 HA19 HA20

4B029 AA07 AA23 BB20 FA15

5B075 ND02 UU18

**THIS PAGE BLANK (USPTO)**